数学小丛书——智慧之花

（3）

乘电梯·翻硬币·游迷宫·下象棋

丁石孙　主编

北京大学出版社

新登字(京)159号

数学小丛书——智慧之花
(3)
乘电梯·翻硬币·游迷宫·下象棋
丁石孙 主编
责任编辑：刘 勇
*
北京大学出版社出版发行
(北京大学校内)
北京大学印刷厂印刷
新华书店经售
*
787×1092毫米 32开本 9.5印张 210千字
1993年6月第一版 1993年6月第一次印刷
印数：0001—4,000册
ISBN 7-301-02085-6/O·314
定价：5.50元

内 容 提 要

本书是北京大学《数学小丛书——智慧之花》的第三本。内容为精选的20个饶有趣味的数学问题。选材生动、有趣，贴近生活，旨在激发中学生和大学生学习数学的兴趣，开拓思想，启迪智慧，使学生得到引人入胜的思维训练。

经常乘电梯的人，等到的电梯是上行的还是下行的可能性是否相同？一摞硬币（正面朝上），按一定规则翻面，直到这摞硬币中的每一个又都是正面朝上为止，共需做多少次翻面？在一个$n\times n$的棋盘中放入n个王，使每行、每列都只有一个，且两两不能相互攻击，有多少种放置方法？怎样按最佳路线游历迷宫？如何用初等数学解决“几何极值问题”？本书对这些并不陌生的问题给出了巧妙、新颖、富有思想、与众不同的回答。为适应数学竞赛学生的需要，本书给出了第32届国际数学奥林匹克竞赛试题与解答。

本书可作为高中学生，中学数学教师和一、二年级大学生的课外读物，对有志参加数学竞赛的学生也有很好的指导意义。本书还可供数学爱好者阅读。

写在前面的话

在一个人所受的基础教育中，数学一直是占着一个特殊地位的，它占用的时间可以说是最多的。也许因为这已是历史上长期以来形成的事实，所以很少有人去作说明，即使有的学生并不喜欢数学，也鼓不起勇气去问个为什么。

数学由于其特殊的形式，给人的印象常常是：一批口诀，一堆公式以及一串定理，但它们在解决生活及其它学科的问题时又是很有用的，于是多数人就硬着头皮按老师教的学下去。这样的理解至多对了一半，因为数学还有另一个方面的重要作用，这就是通过对数学知识的介绍，对数学问题的解决，教会人们一种重要的分析问题，解决问题的思想方法。简单地讲，数学要教会人如何进行逻辑推理，如何进行正确的抽象思维，如何在纷繁的事物中抓住主要的联系，并如何使用明确的概念，等等。

要正确发挥数学课程的教育功能，除去需要教师与学生的积极努力以外，也还需要找到适当的辅助材料和恰当的方法。我们选编这套《数学小丛书——智慧之花》就是为了从这个方面为数学老师（主要是中学的老师）和大学生提供一点帮助，有一部分也可以用作中学生的课外读物。

我们并不认为目前的数学教学大纲的内容太少，太浅，因而要增加或加深教学内容。我们更不想给学生增加习题量以应付考试。恰恰相反，我们认为再向这个方向发展将会造

成极大的危害. 通过我们选择的这些小文章，我们希望能帮助读者对数学有更全面的了解，使大家发现数学不只是“定义、定理、公式、证明”的刻板叙述，而是生动活泼、引人入胜的思维训练. 在这里，读者可以看到如何对各种各样的问题进行精细的分析，又如何逐步把复杂的问题理出头绪，最后给出清晰的答案. 总之，我们希望通过千姿百态的分析与讨论帮助读者了解什么是大家应该从数学学习中学到的思想方法.

我们的目标是这样，但能否达到还有待于实践的检验. 读者读过这些书之后的印象与收获将作出评判. 我们希望大家多提批评意见，帮助我们不断改进我们的工作.

丁石孙

1989年2月

出版说明

现代数学，这个最令人惊叹的智力创造，已经使人类心灵的眼光越过无限的时间，使人类心灵的手延伸到了无边无际的空间。

——N.M.Butler

数学方法渗透进并支配着一切自然科学的“理论”分支。在现代经验科学中，它已越来越成为衡量成就的主要标准。

——J.von Neumann

参与开发一般智力——不是为了今后某一职业的特定需要，应看成是数学教育的基本目标。

——F.Reidt

别把数学想象得那么困难和艰涩，并认为它排斥常识，数学仅仅是常识的一种微妙的形式。

——L.Kelvin

这些著名学者的话表达了我们出版《数学小丛书——智慧之花》的想法和努力的目标。

本丛书的主要对象是：中学数学教师、数学各专业的低年级大学生、部分高中学生以及数学爱好者。所选内容力求生动、有趣，在开始阶段以翻译为主，一年2—3册。

我们希望本丛书能为活跃与推动中学与大学低年级的数学教学、提高中学教师和大学生的数学素质、更好地沟通中

学数学与大学数学以及普及数学知识，做一点有益的工作。

我们水平有限，希望大家多提意见，为了让我们的小花开放得绚丽多姿而共同努力！

《数学小丛书——智慧之花》编委会

1989年2月

目　录

电梯、火车及地铁[①]

A. Wuffle

以下3个疑难问题的数学处理是相同的。

问题1 经常乘电梯的人有这样的体会：除非是在楼的底层或顶层，否则你等来的第一部电梯的运行方向差不多总是与你希望去的方向相反(参见[1,p.10—11])。但是，下面的说法似乎也在理：要下去必须先上来，因此，等到的电梯是上行的还是下行的可能性应该是相同的。那么，这两者为什么可能都是对的呢？[②]

问题2 一个纽约人有两个好朋友，一个住在市中心，另一个住在郊区。他和这两个朋友都很好。因此，当他想去看望他们时，他总是登上在地铁车站赶上的第一列地铁，而不管它是去市中心的还是去郊区的。到两个方向去的地铁班次是一样多的。虽然他无意对这两个好朋友厚此薄彼，但结果是，他去一个朋友处的次数远远超过去另一个朋友处的次数。为什么会这样呢(参见[2])？

问题3([1,p.59—60]) 在美国中西部的一个小镇上住着一位退休的铁路工程师W.Johnson。他工作了大半辈子的

① The pure theory of elevators, *Math. Magazine*, Vol. 55, 1 (1982), 30—37.

② 这里说的电梯不是现在的程控电梯，而是旧式电梯。

那条铁路线正好穿过这个小镇。Johnson患有失眠症，经常会在夜里的某一个奇数钟点（但不固定）醒来，且再也不能入睡。后来，他找到了治疗失眠症的好方法。每当醒来后，他就沿着小镇上那条寂静的街道步行，一直走到与铁路的交叉点。他站在那儿，若有所思地看着表，一直等到有一列火车开来。火车的吼叫声撕破了宁静的夜空，这一情景使这位老铁路心境舒畅。然后他走回家，很快就能入睡。

过了一段时间后，他意外地发现，他所看到的火车大部分是往东开的，只有很少几列是西行的。这位工程师清楚地知道，这条干线上东行火车与西行火车的次数是一样多的，而且它们的交错也很有规律。开始时，他以为一定是自己记错了。于是备了一个小本子，根据他所看到的第一列火车开去的方向，在本子上分别写上“东”或“西”。过了一周，他发现共记了5个“东”且只有2个“西”。接下来的一个星期的记录也几乎如此。他想，是不是因为自己差不多总是在夜间的同一个钟点醒来，因而总是赶上一列东行的火车呢？

被这一情景所困扰，他决定对问题作一个统计研究，并决定白天也进行观察。为客观计，他请一位朋友替他拟了一个长长的随机时刻表，诸如上午9:35，中午12点，下午3:07等等。他按这些钟点准时赶到铁路交叉处，记下看到的第一列火车的运行方向。出乎意料，记录的结果与原先的差不多：在100列火车中，有75列是往东的，只有25列是朝西开的。失望之余，他打电话给附近的大城市的火车站，询问是否有些西行火车已改线了。但回答说“不是”。这一奇怪现象使这位老铁路如此沮丧以致完全失眠，日渐虚弱。

我们先解决第 1 个疑难问题。设一幢大楼有N层且有r部电梯。为简单计，先考虑特殊情形$r=1$。在任意一个时刻，电梯处在$2(N-1)$种状态中的某一种状态，记这些状态分别为$1\uparrow, 2\uparrow, 2\downarrow, 3\uparrow, 3\downarrow, \cdots, (N-1)\uparrow, (N-1)\downarrow, N\downarrow$，其中的箭头表示电梯停在第$i$层之后接下来的运行方向。当然，在顶层和底层，电梯只有一个方向。假设电梯在每两层楼间运行的时间为t，还假设它在每层楼的固定的停留时间也包括在t内(实际上，为简便计，一般将其忽略)。考虑一个高峰时期，假定电梯在每层楼上都要停。如果你正在第k层楼上等电梯，那么你等来的第一部电梯恰好是上行电梯的概率有多大呢？记之为$P(k\uparrow)$。对于$1<k<N$，等来的电梯是上行的，如果它正处在状态$1\uparrow, 2\uparrow, \cdots, (k-1)\uparrow$之一，或处在状态$2\downarrow, 3\downarrow, \cdots, k\downarrow$之一中。这种状况共有$(k-1)+(k-1)=2k-2$种。不妨假定，所有$2N-2$种状态都是等可能的。因此，当$1<k<N$时有

$$P(k\uparrow)=\frac{2(k-1)}{2(N-1)}=\frac{k-1}{N-1} \tag{1}$$

及

$$P(k\downarrow)=\frac{N-k}{N-1}=1-P(k\uparrow). \tag{2}$$

毫无疑问，$P(1\uparrow)=1$，$P(N\downarrow)=1$。因此，举例说来，如果你在加利福尼亚大学社会科学塔 Irvine 的二层楼上（该塔共有 7 层)，当该塔两部电梯中的一部出故障时(这是经常发生的)，你等来的电梯恰好是上行的概率是$\frac{2-1}{7-1}=\frac{1}{6}$。

现在，考虑另一个不同的问题，当然仍与原问题有关。如果你在第k层楼上，则平均需要花多少时间，才能等到第

一部上行电梯呢？考虑下面的一连串状态

$$k\uparrow,(k+1)\uparrow,(k+2)\uparrow,\cdots,(N-1)\uparrow,N\downarrow,(N-1)\downarrow,$$
$$(N-2)\downarrow,(N-3)\downarrow,\cdots,1\uparrow,2\uparrow,3\uparrow,\cdots,(k-1)\uparrow.$$

当然，一共是 $2(N-2)$ 种状态。如果我们是在第 k 层楼上，如果电梯正处在上述状态链中的第一个状态，即 $k\uparrow$，那么我们等到第一部上行电梯的时间或者是 0（正好赶上），或者是 $2(N-2)t$（恰好错过）。不妨设为 $(2N-2)t$；如果电梯正处在第 2 种状态（即 $(k+1)\uparrow$），则等到上行电梯的时间是 $(2N-3)t$；如果电梯正处在第 j 种状态，则等待的时间是 $(2N-j-1)t$。因此，若假定只有一部电梯，则在第 $k(1<k<N)$ 层上的人，等到一部上行电梯的时间平均值 T_1 由公式

$$T_1=\sum_{j=1}^{2N-2}\frac{2N-j-1}{2N-2}t \tag{3}$$

给出。由对称性，(3)式可写成

$$T_1=t\sum_{j=1}^{2N-2}\frac{j}{2N-2}. \tag{4}$$

由熟知的等式 $\sum_{i=1}^{n}i=n(n+1)/2$，(4)式可写成

$$T_1=\frac{(2N-1)(2N-2)t}{2(2N-2)}=\frac{2N-1}{2}t. \tag{5}$$

由上式显然可以看到，不管我们在哪一层上（除了顶层和底层以外），等到一部上行电梯的（平均）时间是常数。同样显然的是，除了在顶层和底层以外，等到一部下行电梯的（平均）时间与等到一部上行电梯的（平均）时间相同。但是，正如前面已经证明的那样，等到的第一部电梯是上行电梯的概率与

在哪一层有关，也与楼共有多少层有关，见(1)和(2)．(如果假定你总能及时赶上一部停在你所在楼层的电梯，但不一定开往你想去的方向，则 T_1 应为

$$T_1=\frac{2N-3}{2}t.$$

另外，还可将某些概率考虑进去，例如，错过停在你所在的楼层上且又是你要乘的方向的电梯的概率，进而修改上述公式。但是，本文将不讨论这种复杂情形．)

现在分析第 2 个疑难问题．类似于上面的讨论可找到问题的症结。假设地铁每小时一班．去市中心的地铁在每小时的第50分到达，而去郊区的地铁正钟点到达．于是很容易看到，这位纽约人去市中心的次数将是去郊区次数的几乎 5 倍。因为只有在去市中心的地铁刚开走，而去郊区的地铁还未到达的10分钟里他到达车站，才能乘车到郊区去（显然，如果去市中心的地铁在每小时的第30分到达，则他去市中心和郊区的次数会差不多）。另外，无论是去市中心还是去郊区，他等车时间的期望值都是半小时。

第 3 个疑难问题本质上与第 2 个疑难问题相同．假定火车按固定的时刻表(不妨设每12小时一趟)分别从 2 个终点站开出。那个中西部小镇离开西面的火车终点站(洛杉矶)与离开东面的火车站(芝加哥)的距离不等，使得那个老铁路看到的第一列火车是东行的次数远远超过西行的火车。

这 3 个疑难问题解决后，我们继续来研究一下电梯及等电梯的人．这部分讨论可以看作是对所谓“电梯疯狂”现象的研究，这一名词是由那些总是等不上想去方向的电梯的人创造的。

设 $P_k(U)$ 是一个在第 k 层上的人能乘上一部上行电梯的条件概率，$P_k(D)$ 表示他能乘上一部下行电梯的条件概率。首先假定，除了高峰时间外，大楼的每一层对电梯的需求是相同的，即

$$P_k(U)=\frac{N-k}{N-1}$$

及

$$P_k(D)=\frac{k-1}{N-1}.$$

这种假定对于独家公司占有的大楼并不是不合理的。

对某个在第 k 层上想乘电梯的人，定义失望指标 f_k 为等来的第一部电梯的方向正好与他的愿望相反的概率。当 $0<k<N$ 时有(仍假设只有一部电梯)

$$f_k=P_k(U)P(k\downarrow)+P_k(D)P(k\uparrow)$$

$$=\left(\frac{N-k}{N-1}\right)\left(\frac{N-k}{N-1}\right)+\left(\frac{k-1}{N-1}\right)\left(\frac{k-1}{N-1}\right), \quad (6)$$

$$1<k<N.$$

化简后得

$$f_k=1-\frac{2(N-k)(k-1)}{(N-1)^2}. \quad (7)$$

由(7)式显然可得

$$f_k=f_{N-k+1}.$$

当 $k>N/2$ 时，有 $(N-k)(k-1)>(N-k-1)k$。因此由(7)式容易看到，若 N 是奇数，则当 $k=(N+1)/2$ 时，f_k 取最小值。事实上，在只有一部电梯的大楼里，你若越靠近顶层或底层(除了在顶层或底层外)，则你受到"电梯疯狂"现

象的折磨越厉害。另外，由(7)可直接算得 $f_{(N+1)/2}=1/2$。

当你在楼的顶层或底层时，等来的电梯总是开往你想去的方向的，因此可以定义 $f_1=f_N=0$。

根据上述计算，在一部电梯的情形时，失望指标的期望值（记为 F_1）是

$$F_1=\sum_{k=1}^{N} f_k \cdot p(k), \tag{8}$$

其中 $p(k)$ 表示在 k 层楼上想乘电梯的人占全楼想乘电梯的人的比率。如果假定 $p(k)=1/N$ 并继续假设电梯在每一层楼停的可能性相同，由（7）得

$$F_1=\frac{1}{N}\sum_{k=2}^{N-1}\frac{1}{(N-1)^2}[2k^2-2k(N+1)+N^2+1]$$

$$=\frac{(N-2)(2N-3)}{3N(N-1)}.$$

虽然对于小的 $N(N<10)$ 有 $F_1<1/2$，但当 N 充分大时，显然有 F_1 接近 2/3。

现在分析在上下班高峰时的情形，仍先假定只有一部电梯。在早晨的高峰时刻，每个人都想乘电梯上行，而在下午的高峰时刻，每个人又都想乘电梯下行。在早晨时，因为大家都在一楼，因此失望值是 0 。但下午的情形不同了。如果同上文一样，假定人群是均匀分布在每层楼上的，即 $p(k)=1/N$，则当 $k<N$ 时有

$$f_k=\frac{k-1}{N-1} \tag{9}$$

及

$$F_1=\sum_{k=1}^{N}\frac{k-1}{N(N-1)}=\frac{(N-1)(N-2)}{2N(N-1)}=\frac{N-2}{2N}. \tag{10}$$

由简单的计算可以证明：当 $k>(N+1)/2$ 时，等电梯的“正常”失望指标（由(6)给出）小于下午高峰时刻的失望指标（由(9)给出）；当 $k<(N+1)/2$ 时，相反的结论成立。**因此，在一部电梯的情形，越在楼的高层，越接近下午的高峰时刻，“电梯疯狂”越接近它的最强点**（我们忽略不计诸如疲倦性及急于回家等复杂因素，它们也可以被假定在下午的晚些时候达到最大值）。但是，当N充分大时，“正常”时刻的 F_1 接近 2/3，而高峰时刻的 F_1 接近 1/2（见表达式(10)）。因此，在我们作出的简化假定下，高峰时刻的失望指标（依等来的电梯是相反方向的计）的期望值小于“正常”时刻的。

绝大多数高层建筑都配备多部电梯，因此，我们接下来研究电梯数 $r\geqslant 2$ 的情形，仍然设电梯在每一层停的可能性都相同。

先考虑 $r=2$。设某人在第 k 层上等电梯。令 $q_1^{(k)}$是电梯1作为下（上）行电梯到达第 k 层所需经过的楼层数，类似地定义 $q_2^{(k)}$。显然，只有等电梯运行了 $\min(q_1^{(k)},q_2^{(k)})$ 层之后，才能在第 k 层等到一部下（上）行电梯。

不难看到，事件“$\min(q_1^{(k)},q_2^{(k)})=h$”的概率与 k 无关，因此可以去掉上角标（k）。由概率的加法公式及条件概率得

$$\begin{aligned}P(\min(q_1,q_2)=h)&=P(q_1=h\,|\,q_2>h)P(q_2>h)\\&\quad+P(q_1>h\,|\,q_2=h)P(q_2=h)\\&\quad+P(q_1=h\,|\,q_2=h)P(q_2=h).\end{aligned}$$

代入计算结果得

$$P(\min(q_1,q_2)=h)=\frac{1}{2(N-1)}$$

$$\times\left[\frac{2N-2-h}{2(N-1)}+\frac{2N-2-h}{2(N-1)}+\frac{1}{2(N-1)}\right]$$

$$=\frac{4N-3-2h}{4(N-1)^2}.$$

设电梯在每相邻两层间运行的时间为 t 并忽略不计在每层停的时间，则在有两部电梯的情形时，等一步下（上）行电梯的时间期望值 T_2 是

$$T_2=\sum_{h=1}^{2N-2}\left(\frac{4N-3-2h}{4(N-1)^2}\right)ht=\frac{(2N-1)(4N-3)t}{12(N-1)}. \qquad (11)$$

当电梯数增大时，上述计算显得太笨拙。有一种简单得多且易于推广的方法可用于得到相同的结果。我们可以看到

$$P(\min(q_1,q_2)\geqslant h)=\left[\frac{2(N-1)-h+1}{2(N-1)}\right]^2,$$

因此

$$T_2=\sum_{h=1}^{2N-2}P(\min(q_1,q_2)\geqslant h)t$$

$$=\sum_{h=1}^{2N-2}\left[\frac{2(N-1)-h+1}{2(N-1)}\right]^2 t$$

$$=\frac{(2N-1)(4N-3)t}{12(N-1)}.$$

更一般地，在有 r 部电梯的情形可得

$$P(\min(q_1,q_2,\cdots,q_r)\geqslant h)=\left[\frac{2(N-1)-h+1}{2(N-1)}\right]^r$$

及

$$T_r=\sum_{h=1}^{2N-2}P(\min(q_1,q_2,\cdots,q_r)\geqslant h)t$$

$$=\sum_{h=1}^{2N-2}\left[\frac{2(N-1)-h+1}{2(N-1)}\right]^r t$$

$$=\sum_{h=1}^{2N-2}\frac{h^r}{(2(N-1))^r}t.$$

可以用好几个公式计算$\sum h^r$（见［3］）．有一种比较简单的方法是假定N很大，可得下面的很有用的结果：

$$T_r\approx Nt\int_0^2\left(\frac{x}{2}\right)^r dx=\frac{2Nt}{r+1}. \qquad (12)$$

注意，当$r=1$时，由（12）式得$T_1\approx Nt$，这与我们在前面算得的离散值表达式$T_1=\left(N-\frac{1}{2}\right)t$相差不大；当$r=2$时，由（12）式得$T_2\approx 2Nt/3$，当$N$充分大时，由（11）式也可得同样的结果．正如经验告诉我们的那样：电梯越多，等下一部上（下）行电梯的时间的期望值就越小．

当$r=2$时，可以得到计算$P_2(k\uparrow)$及$P_2(k\downarrow)$的公式，其中$P_2(k\uparrow)$和$P_2(k\downarrow)$表示到达k层的下一部电梯分别是上行和下行的概率．因为有两部电梯，所以有一个概率结，即上行电梯和下行电梯同时到达的概率，记为$P_2(k\updownarrow)$．我们用概率树来解决这一问题，见图1（设在大楼的第k层等电梯，而大楼配备两部电梯．符号“>”表示“先于…到．图中下半部分的概率是关于情形$k\geqslant(N+1)/2$的)．

在图1中，$D_{ki}(U_{ki})$表示第i部电梯到达第k层时是下（上）行的，符号“>”表示“在…之前到达”．例如，D_{k1}

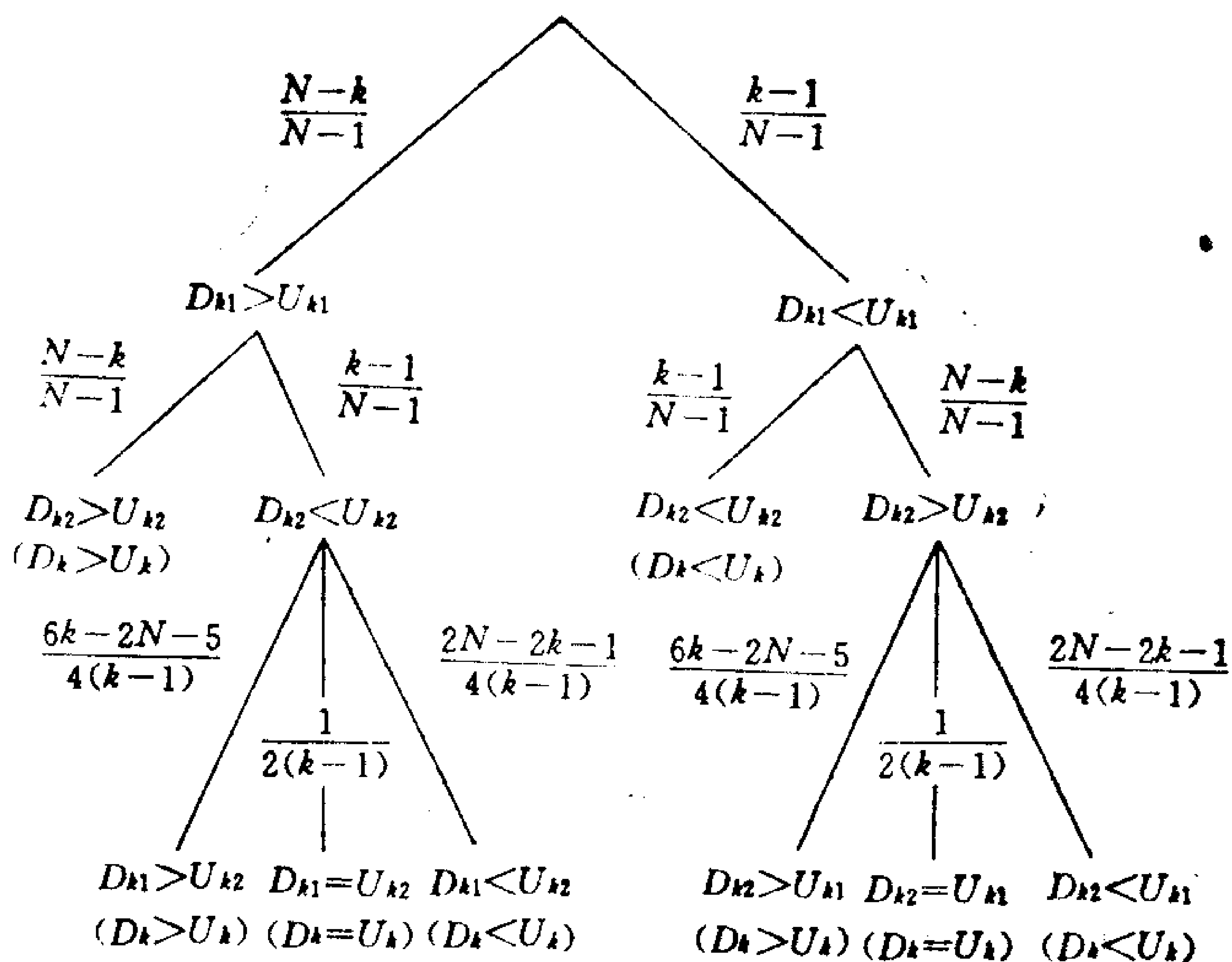


图 1　下一部电梯是上行电梯和下行电梯的概率

$>U_{ki}$表示事件“第 i 部电梯到达第 k 层时是下行而不是上行的”。$D_{ki}<U_{kj}$及$D_{ki}=U_{kj}$也类似地定义。结论标在概率树的分枝端点上，例如，如果 $D_{k1}>U_{k1}$及$D_{k2}>U_{k2}$，则到达第 k 层的下一部电梯必定是下行的，在图 1 中记为 $(D_k>U_k)$。概率树上标出的概率可以直接计算出来。我们只计算一种典型情形作为例子：其中一部电梯下次开到第 k 层时是下行的，而另一部电梯下次开到第 k 层时是上行的。这时需要计算许多条件概率。我们以计算 $P(D_{k2}>U_{k1}|D_{k1}<U_{k1}\wedge D_{k2}>U_{k2})$ 为例。由 $D_{k2}>U_{k2}$，得知第 2 部电梯正在第 k 层以上（包括刚离开第 k 层），而且最多运行$2(N-k)$层后将作为下行电梯到达第 k 层。由$D_{k1}<U_{k1}$，

得知第 1 部电梯正在第 k 层以下（包括刚离开第 k 层），而且最多运行$2(k-1)$ 层后，将作为上行电梯到达第 k 层。因此，对于 $k\geqslant(N+1)/2$，有

$$P(D_{k2}>U_{k1}\,|\,D_{k1}<U_{k1}\wedge D_{k2}>U_{k2})$$

$$=\sum_{k=1}^{2(N-k)}P(q_1>h\,|\,q_2=h)P(q_2=h)$$

$$=\sum_{h=1}^{2(N-k)}\left[\left(\frac{2(k-1)-h}{2(k-1)}\right)\left(\frac{1}{2(N-k)}\right)\right]$$

$$=\frac{6k-2N-5}{4(k-1)}.$$

图 1 中标出的概率值可用类似的方法算出来。将图 1 中的有关概率相加可得：对于 $k\geqslant(N+1)/2$，有

$$P_2(k\uparrow)=P(D_k<U_k)$$

$$=\frac{N-k}{N-1}\cdot\frac{k-1}{N-1}\cdot\frac{2N-2k-1}{4(k-1)}+\frac{k-1}{N-1}\cdot\frac{k-1}{N-1}$$

$$+\frac{k-1}{N-1}\cdot\frac{N-k}{N-1}\cdot\frac{2N-2k-1}{4(k-1)}$$

$$=\frac{2(k-1)^2+(N-k)(2N-2k-1)}{2(N-1)^2}; \qquad (13)$$

$$P_2(k\downarrow)=P(D_k>U_k)$$

$$=\frac{N-k}{N-1}\cdot\frac{N-k}{N-1}+\frac{N-k}{N-1}\cdot\frac{k-1}{N-1}\cdot\frac{6k-2N-5}{4(k-1)}$$

$$+\frac{k-1}{N-1}\cdot\frac{N-k}{N-1}\cdot\frac{6k-2N-5}{4(k-1)}$$

$$=\frac{(N-k)(4k-5)}{2(N-1)^2}, \qquad (14)$$

而在概率结上，有

$$P_2(k\updownarrow)=P(D_k=U_k)$$

$$=\frac{N-k}{N-1}\cdot\frac{k-1}{N-1}\cdot\frac{1}{2(k-1)}+\frac{k-1}{N-1}\cdot\frac{N-k}{N-1}\cdot\frac{1}{2(k-1)}$$

$$=\frac{N-k}{(N-1)^2}.$$

对于$k<(N+1)/2$，只需将等式（13）和（14）右边的项互换一下即可。当然，对于$k=(N+1)/2$，这两个值是相等的。

另外，我们还可以定义失望指标f_k为：

$$f_k=P_k(U)P_2(k\downarrow)+P_k(D)P_2(k\uparrow)$$

$$=\frac{(N-k)^2(4k-5)}{2(N-1)^3}$$

$$+\frac{(k-1)[2(k-1)^2+(N-k)(2N-2k-1)]}{2(N-1)^3}.$$

将等式(13),(14)与等式(1),(2)相比较得：当$k>(N+1)/2$时有 $P_2(k\uparrow)>P(k\uparrow)$ 及 $P_2(k\downarrow)<P(k\downarrow)$，而当$k<(N+1)/2$时,反向不等式成立.这究竟意谓着什么呢？它告诉我们：增加了第2部电梯后，仅仅对一半层次（下面的一半楼层）减少了下一部电梯是上（下）行的可能性。**增加电梯趋向于去"拉平"到达楼层的电梯是上行还是下行的概率。**

对相当大的r及N，下面的近似值公式是有用的：

$$P(k\uparrow)=P(k\downarrow)\approx\frac{2N-3}{4(N-1)}\approx\frac{1}{2},$$

$$P(k\downarrow)=\frac{1}{2(N-1)}.$$

因此，对充分大的 r 及 N 有

$$f_k\approx\frac{1}{2},\quad F_r\approx\frac{1}{2},$$

其中 F_r 是失望指标期望值（见(8)）。

电梯、火车及地铁都是封闭的环状系统。本文讨论的这几个疑难问题阐述了"频数"及"状态"之间的联系(见[1])。有兴趣的读者还可以考虑其他的封闭环状系统并讨论其中提出的数学问题。例如，纽约市的地铁既有快车又有慢车（快车是因为停站少），你还可以列出一些条件，然后考虑赶上错误方向的地铁的问题。（提示：你只能在只有慢车停靠的车站候车吗？）

参 考 文 献

[1] G.Gamow and M.Stern,Puzzle-Math.Viking,New York, 1958.

[2] F.Mosteller,Fifty challenging problems in probability (with solutions),Addison-Wesley, Reading, MA, 1965.

[3] J.W.Paul,On the sum of the products of the first integers, *Amer.Math.Monthly*, 78(1971), 271—272.

（朱学贤编译，潘承彪校）

翻 硬 币[1]

B.B.Newman

一摞硬币共M枚，每枚都是正面朝上(本文中简记为H，背面朝上简记为T——译者注)。取下最上面的1枚硬币，将它翻面后放回原处。然后取下最上面的2枚硬币，将它们一起翻面后再放回原处。再取3枚、取4枚、…，直至整摞硬币都按上述方法处理过。然后再从这摞硬币最上面的1枚开始，重复刚才的做法。这样一直做下去，直到这摞硬币中的每一个又都是正面朝上为止。

问题：这种情形是否一定出现？如果出现，则一共需要做多少次翻面？

现以$M=4$为例，整个过程如下（其中的括号表示每一次被翻面的硬币）：

1	H	T	T	T	T	H	H	T	T	H	T	H
2	H	H	H	T	T	T	T	H	H	H	T	H
3	H	H	H	H	H	H	H	T	T	T	T	H
4	H	H	H	H	H	H	H	H	H	H	H	H

因此，4个一摞的硬币共需做11次翻面。这一问题是由Manchester大学科学技术研究所的John Gilder及Iain Bride构想的。在给作者的一封信中，John Gilder写道：“它的历

[1] The Flippin'Coins Problem, *Math.Magazine*, 54(1981), 51—59.译者作了若干改动.

史比较短。在边等公共交通边要弄我的那些小钢镚时，构想了这一问题。”最初的提法略有不同，目前的形式是由他的同事 Iain Bride 给出的。对特殊的M解出这一问题，可以作为初等计算机程序课中的一个很具启发性的习题，Graham Birtwistle 及其他一些人将它吸收进讲授 Simula 语言的书中（[1]）。对不同的M计算出的翻面次数见表 1 。鉴于其中的结果，Birtwistle 提出下列猜想。

表 1

M（硬币个数）	翻面次数	M（硬币个数）	翻面次数
1	2	17	204
2	3	18	323
3	9	19	228
4	11	20	199
5	24	21	146
6	35	22	264
7	28	23	529
8	31	24	504
9	80	25	200
10	60	26	675
11	121	27	540
12	119	28	251
13	116	29	840
14	195	30	899
15	75	31	186
16	79	32	191

猜想1　一摞M个硬币的翻面次数是 Mk 或 $Mk-1$。

猜想2　翻面次数不大于 $M^2(M>1)$。

猜想3　对于$2^n \leqslant M \leqslant 2^{n+1}-1$，当$M=2^n$时有最少的翻

面次数，是$M(n+1)-1$。若$M=2^{n+1}-1$，也有最少的翻面次数，是 $M(n+2)$。

本文证明这 3 个猜想都是正确的。

在一摞硬币中，与每枚硬币相联系的是它所处的位置及它的状态（正面朝上还是背面朝上）。每一次翻面既改变硬币在摞中的位置，又使所取一迭硬币的状态发生变化。我们在下文中将看到，当整摞硬币都被翻面之后，知道每枚硬币在摞中的位置，是证明猜想的钥匙。表面看来，只考虑每一次翻面之后硬币位置的置换将会得到一个不完整的分析，因为在每一个中间环节上整摞中的一部分也许组成了正面朝上的一迭。虽然硬币位置的那些置换是证明猜想的关键，但我们最终关心的并不是每枚硬币在摞中所处的位置，而是它究竟是正面朝上还是背面朝上？见图 1 。图 1 给出的是一摞 4 枚不同大小的硬币，在每一次翻面之后每枚硬币的位置及状态（白色表示H，黑色表示T）的变化。作11次翻面之后，所有硬币都正面朝上，但位置恰好倒过来。

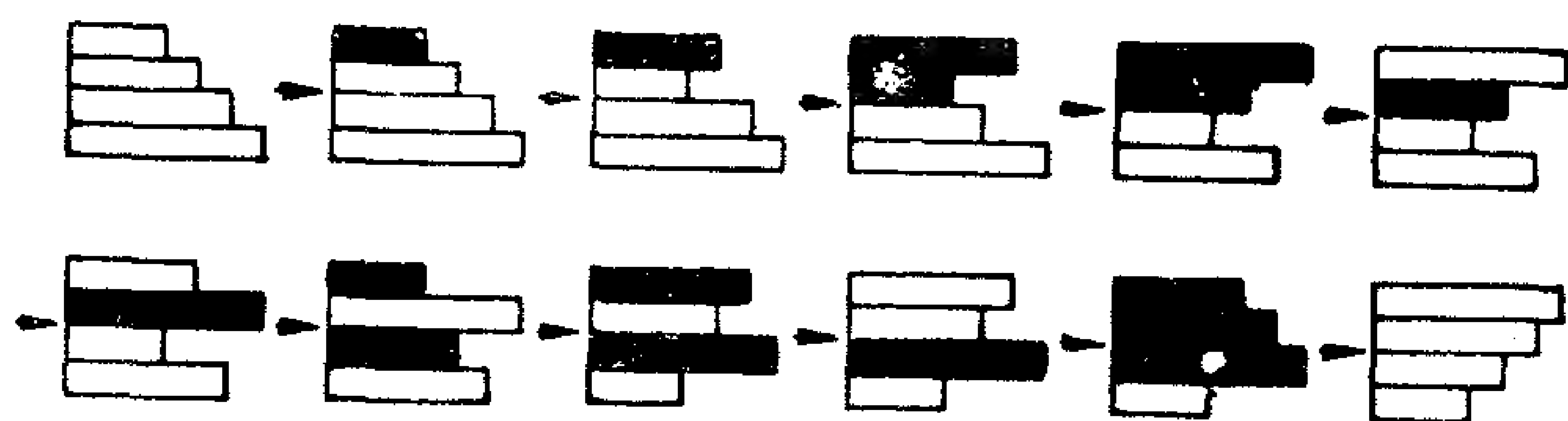


图 1

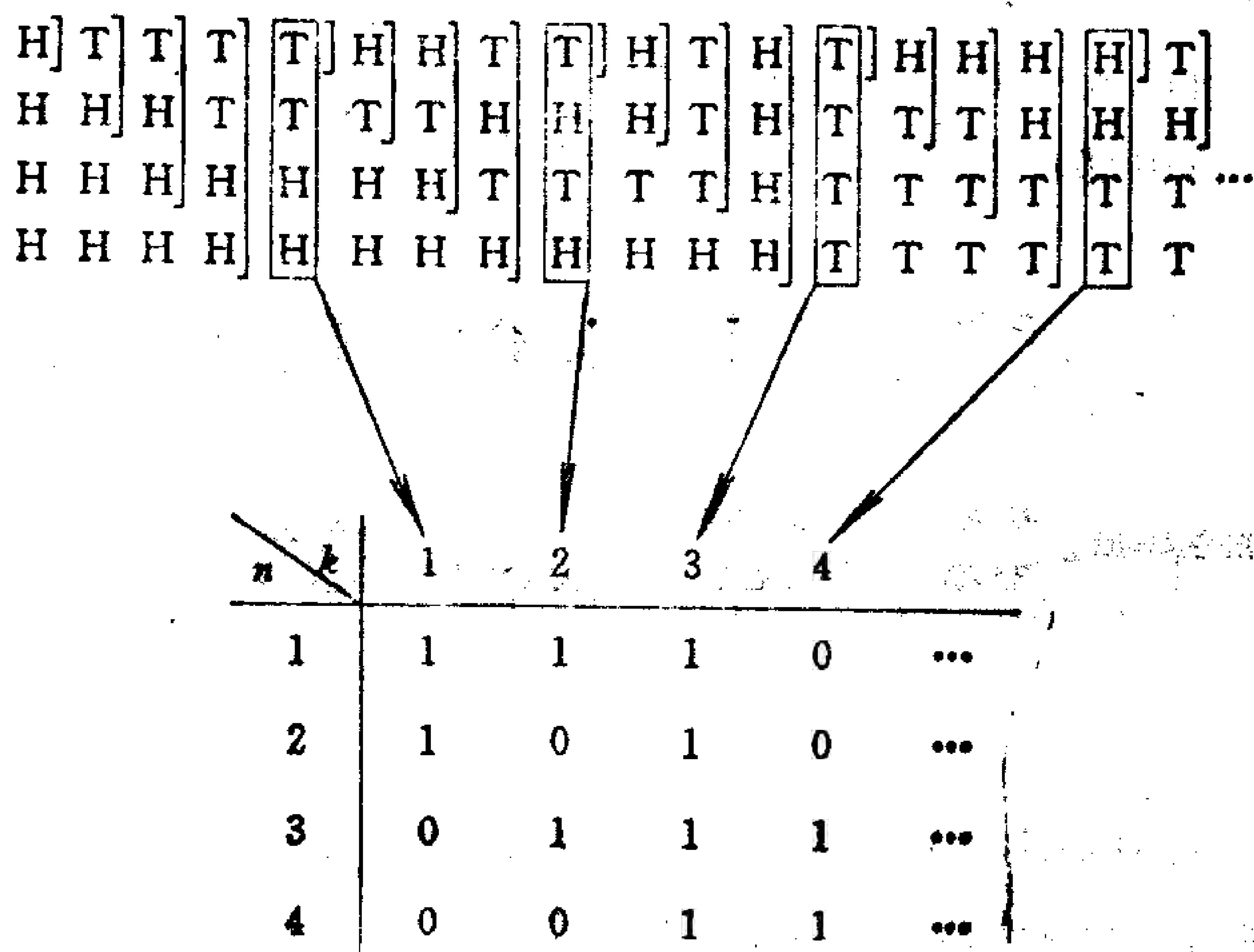
下面的引理 1 实质上是硬币翻面的一个饶有趣味及有用的刻画。单独用它即可证明猜想 1 ，附加上一点点初等群论，可以证明猜想 2 。我们先假定引理 1 成立，用它证明猜

想 1 和猜想 2，然后深入到讨论置换，最后证明猜想 3 及引理 1 。

这里提及的置换具有一些很有趣味的性质，它出现在若干个研究课题中，这些课题可追溯到1773年（[3]）。本文中增加的新东西是将它和硬币的翻面联系起来。所用的技巧可以从硬币翻面的 2 种状态推广到多种状态，用于解决诸如下面的问题：一摞M个表（都在同一个钟点），每一个表每翻一次面就慢 n 个小时，问需翻面多少次才能使所有的表又都指示同一个钟点？

猜想1及猜想2的证明

如果猜想 1 成立，则一摞硬币全部是正面朝上的情况，仅当在某一次将整摞M个硬币翻面时出现，或者恰好在这之前



n \ k	1	2	3	4	
1	1	1	1	0	…
2	1	0	1	0	…
3	0	1	1	1	…
4	0	0	1	1	…

图2　（表中 k 表示第 k 次整摞翻面）

一次出现。因而我们只需观察每一次将整摞硬币翻面之后摞中硬币的状态。即，我们感兴趣的是第M次、$2M$次、$3M$次、…翻面之后摞中硬币的状态。为此，抽出这种整摞翻面后的结果列，将它们放在一起，并用0表示正面朝上，用1表示背面朝上。对于$M=4$可得图2。

我们将视这种抽取后结果的第n行为二进小数$p_M(n)$。对于$M=4$，表2给出$p_4(n)$（$n=1,2,3,4$）的二进小数表

表 2

$p_4(1)=0.1110\cdots$
$p_4(2)=0.1010\cdots$
$p_4(3)=0.0111\cdots$
$p_4(4)=0.0011\cdots$

示。从硬币摞中摘取出这些数之后，可以看到$p_M(n)$必定是循环小数，因而表示有理数（这是因为，若假定猜想1成立，则对每一个M，经过有限次整摞翻面之后，所有的硬币又都重新成为正面朝上。因此，二进小数$p_M(n)$中必定有一个重复出现的数字段）。然而，这些小数的有理数表示的意料不到的简单性是证明猜想的关键。

引理1 $p_M(n)=\dfrac{2M+2-2n}{2M+1}$.

在本文的最后一节中将证明此引理。现在先假定引理1成立并使用之。这一引理不仅仅用于分析整摞翻面的结果，而且还提供了观察硬币摞结构的一个有力工具，对于$M=4$，

用引理1计算的结果见表3(注意，为得到二进小数的有理表示，需计算相应的几何级数的和。例如，$.\overline{111000}=\sum_{n=0}^{\infty}(2^{-1}+2^{-2}+2^{-3})2^{-6n}$)。这些有理数给出了表2中出现在

表 3

n	$p_M(n)$	$\frac{2M+2-2n}{2M+1}$
1	$.\overline{111000}$	8/9
2	$.\overline{10}1010$	6/9
3	$.\overline{011100}$	4/9
4	$.\overline{001110}$	2/9

每一行中的数字的一个简单表示，但更重要的是，可以知道出现在每一列中的数字。这是很容易得到的，因为，如果$p_M(n)\geqslant 1/2$，则'1'将出现在小数点后的第1个位置上；如果$p_n(M)<1/2$，则0将出现在这个位置上。因此，相应于表3中的数字8/9,6/9,4/9及2/9，第1个二进数字列是1,1,0,0。为了得到表2的第k列数字（即相应于第k次整摞翻面），可以用类似的推理。先用2^{k-1}乘以二进小数从而将小数点移到所需位置，然后再来看所得数的小数部分。类似地，若小数部分不小于1/2，则得到数字'1'，否则得数字'0'。然而，在使用有理表示时可以看到，这种二进位制的移动小数点后取小数部分，可以按通常的有理数乘法来实现：先用2^{k-1}乘$(2M+2-2n)/(2M+1)$，然后依$\bmod(2M+1)$约简分子所得到的分数。(为使读者易于理解，我们简述"依

mod(2M+1) 约简”如下：

设 a 与 b 都是正整数且 $a \geqslant 1$。由带余除法知，存在整数 q 及 r 使

$$b = qa + r, \quad 0 \leqslant r < a。$$

我们称 r 是 b 对模 a 的最小非负剩余，记为 $b(\bmod(a))$。从而，上述取小数部分的过程即为：小数部分就等于

$$2^{k-1} \cdot \frac{2M+2-2n}{2M+1} = \frac{(M+1-n)2^k}{2M+1}$$

的小数部分，即分数

$$\frac{((M+1-n)2^k)(\bmod(2M+1))}{2M+1},$$

显见，这分数的分子就是 $(M+1-n)2^k$ 对模 $2M+1$ 的最小非负剩余.)

当 $m=4$ 时，与 $p_M(n)$ 的第 k 个数字相应的这种分数见表4．显见，在第 k 列中的分数的分子，从上到下是由下述

表 4

n	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
1	8/9	7/9	5/9	1/9	2/9	4/9
2	6/9	3/9	6/9	3/9	6/9	3/9
3	4/9	8/9	7/9	5/9	1/9	2/9
4	2/9	4/9	8/9	7/9	5/9	1/9

等差数列对模 $2M+1$ 的最小非负剩余组成的：

$$M \times 2^k, (M-1) \times 2^k, \cdots, 3 \times 2^k, 2 \times 2^k, 1 \times 2^k.$$

只要分子 $(M+1-n)\times 2^k(\bmod(2M+1))$ 小于 $M+1$，则0将在 $p_M(n)$ 的相应位置上。为使第 k 列上的数字全为0，必须使所有分子 $(M+1-n)\times 2^k(\bmod(2M+1))$均小于 $M+1$，$n=1,\cdots,M$。容易看到仅当 $2^k\equiv 1(\bmod(2M+1))$ 时才有这种情形出现（为什么?），此时分子序列 $(\bmod(2M+1))$为

$$M, M-1, \cdots, 3, 2, 1.$$

同样地，容易看到仅当 $2^k\equiv 2M\equiv -1(\bmod(2M+1))$ 时，第 k 列上的数字才皆为1（为什么?），此时分子序列 $(\bmod(2M+1))$为

$$M+1, M+2, \cdots, 2M-2, 2M-1, 2M.$$

如果用原先的硬币摞来解释上述有关二进数字的讨论，可以这样说，所有硬币正面朝上的情形在 k 次整摞（M个硬币）翻面之后马上出现，其中 $2^k\equiv 1(\bmod(2M+1))$，或者，在 k 次整摞翻面之前出现，此时 $2^k\equiv -1(\bmod(2M+1))$。

上述讨论并没有完全证实猜想 1，因为仍然存在这种可能性：所有硬币正面朝上的情形在下一个整摞翻面之前二次或二次以上时出现。这只可能在这种情形发生：在某一次整摞翻面后，其底部的硬币(至少 2 枚(为什么?)) 全是正面朝上，而其余的硬币按一枚正面朝上一枚背面朝上的次序（或者反过来）排列（为什么?）。作为一种假想的情形，考虑 $m=8$，一次整摞翻面的结果如表 5 中的第 1 列。则经过 5 次翻面后，全部硬币都将是正面朝上的。我们将证明，在一次整摞翻面后，表 5 中的第 1 列那样的情形（对任意M）是不可能发生的。注意，在一次整摞翻面后，如果只有 1 枚硬币是背面朝上的而且是在摞顶，那么这将意味着这摞硬币在这之前 2 次翻面时就已是全部正面朝上了。在一次整摞翻面

表5 ($M=8$)

T	H	T	H	T	H
H	H	T	H	T	H
T	T	T	H	T	H
H	H	H	H	T	H
T	T	T	T	T	H
H	H	H	H	H	H
H	H	H	H	H	H
H	H	H	H	H	H

后如表5中第1列所示的硬币状态列，相应于数列 $(\bmod(2M+1))$.

$$Ma, (M-1)a, \cdots, 3a, 2a, a。$$

假定摞底的 $r-1$ 枚硬币都是正面朝上的（由于至少有2枚，因此 $r\geqslant 3$），而从摞底数起第 r 枚硬币是背面朝上的。这时有

$$la(\bmod(2M+1))\leqslant M; \quad 1\leqslant l\leqslant r-1,$$

及

$$ra(\bmod(2M+1))>M.$$

由此易得（注意 $r\geqslant 3$）

$$ra(\bmod(2M+1))=(r-1)a(\bmod(2M+1))+a$$
$$\leqslant M+M=2M;$$

$$(r+1)a(\bmod(2M+1))$$
$$=(r-1)a(\bmod(2M+1))+2a(\bmod(2M+1))$$
$$\leqslant M+M=2M.$$

由此及等式

$$(2a)(\bmod(2M+1))=2(a(\bmod(2M+1))$$

即得

$$(r+1)a(\bmod(2M+1))\geqslant ra(\bmod(2M+1))>M,$$

从而证明了一枚正面朝上一枚背面朝上交错排列的情形是不可能发生的。因而我们证明了，整摞硬币都是正面朝上的充分必要条件是作了 Mk 次翻面，其中 $2^k\equiv1(\bmod(2M+1))$，或者作了 $Mk-1$ 次翻面，其中 $2^k\equiv-1(\bmod(2M+1))$。猜想 1 得证①。

有意思的问题是去得到可能的 M 值，使得存在整数 k，$2^k\equiv-1(\bmod(2M+1))$，因为此时翻面次数不是 M 的整数倍。如果 k 是偶数，则 -1 是 $2M+1$ 的二次剩余（设 $M\geqslant1$，a 是整数。若存在整数 x 使 $x^2\equiv a(\bmod M)$，则称 a 是模 M 的二次剩余，否则称 a 是模 M 的二次非剩余），因而也是每个素数除以 $2M+1$ 的二次剩余。如果 k 是奇数，则 -2 是 $2M+1$ 的二次剩余，因而也是每个素数除以 $2M+1$ 的二次剩余。下面两个数论中的结果（见[2]，p.135 及 p.139，或任何一本初等数论的教科书）是这里要用到的。

引理2 数 -1 是所有形式为 $8n+1$ 或 $8n+5$ 的素数的二次剩余，但它是所有形式为 $8n+3$ 或 $8n+7$ 的素数的二次非剩余。

引理3 数 -2 是所有形式为 $8n+1$ 或 $8n+3$ 的素数的二次剩余，但它是所有形式为 $8n+5$ 或 $8n+7$ 的素数的二次非剩余。

由这两条引理可得：如果 p 是一个素数，整除 $2M+1$，那么，其形式为：

(i) $8n+7$ 时，则 $2^k\equiv-1(\bmod(2M+1))$ 无解；

(ii) $8n+5$ 时，则 $2^k\equiv-1(\bmod(2M+1))$ 仅当 k 是偶数

---

① 下面将证明这样的 k 一定存在。

时有解；

(iii) $8n+3$ 时，则 $2^k\equiv-1(\mathrm{mod}(2M+1))$ 仅当 k 为奇数时有解。

因而，如果 $2M+1$ 能被形式为 $8n+7$ 的任一素数整除，或能被形式分别为 $8n+3$ 及 $8n+5$ 的任意两个素数的乘积整除，则不存在整数 k 使 $2^k\equiv-1(\mathrm{mod}(2M+1))$. 形式分别为 $8n+3$ 及 $8n+5$ 的两个素数的乘积是形式为 $8n+7$ 的数。同样，$8n+7$ 也一定有素数因子其形式为(a) $8n+7$ 或，(b) $8n+3$ 及 $8n+5$. 因此，我们可以结合这两种情形得：如果 $2M+1$ 能被形式为 $8n+7$ 的任意数整除，则 $2^k\not\equiv-1(\mathrm{mod}(2M+1))$。

例如，如果 M 是形式为 $4K-1$, $7K-4$ 或 $12K-2$ 的数，则 $2M+1$ 分别是形式 $8K-1$, $14K-7$ 及 $24K-3$ 的数，它们都能被形式为 $8n+7$ 的数整除。对这些 M，翻面次数都是 M 的倍数。这几个形式解释了表 1 中那些 M 整除翻面次数的除了 $M=25$ 以外的所有情形。现在可以证实猜想 2 了。由上面得到的结果，为证明 M 个硬币一摞的翻面次数至多是 M^2 次，只需证明：存在整数 $k\leqslant M$ 使得 $2^k\equiv\pm1(\mathrm{mod}(2M+1))$. 为此目的引进初等群论的有关结果。

考虑小于 $2M+1$ 且与 $2M+1$ 互素的全体正整数所组成的集合 G，将模 $2M+1$ 的乘法作为 G 中的二元运算 $\odot$，则 G 构成一个群①. G 的阶必定是某个偶整数 $2n$，因为如果 x 与 $2M+1$ 互素，则差 $2M+1-x$ 也与 $2M+1$ 互素，从而群中的元素可以按对分开。显然有 $2n\leqslant 2M$。我们感兴趣的是由 2 生成的 G 的子群。有两种可能出现的情况，以 $M=4$ 及

---

① $\forall a,b\in G$，定义 $a\odot b=(a\cdot b)(\mathrm{mod}(2M+1))$。请读者证明 G 是一个群。

$M=7$ 为例。

$M=4$：$G=\{1,2,4,5,7,8\}$及 2 生成 G。

$M=7$：$G=\{1,2,4,7,8,11,13,14\}$及 2 生成真子群$\{2, 2^2=4, 2^3=8, 2^4=1\}$。

如果 2 的方幂生成整个群 G（如同 $M=4$ 的情形），则有 $2^{2n}-1\equiv 0(\bmod(2M+1))$或者 $(2^n+1)(2^n-1)\equiv 0(\bmod(2M+1))$。因为 $2n$ 是 G 的阶，$2^n\not\equiv 1$，因此 $2^n\equiv -1(\bmod(2M+1))$。于是存在整数 $k(=n)\leqslant M$ 具有所需的性质。

如果 2 的方幂生成 G 的一个真子群（如同 $M=7$ 的情形），则由 Lagrange 定理，此子群的阶 k 是 G 的阶的一个因子。因为 $2^k\equiv 1(\bmod(2M+1))$ 且 k 必定满足不等式 $k\leqslant\frac{1}{2}\cdot 2n\leqslant M$，所以我们证明了存在整数 k 具有所需的性质。

猜想 3 及引理 1 的证明

到目前为止，我们关心的仅仅是硬币的哪个面朝上。为证明猜想 3 及引理 1，需要检查一下硬币在摞中位置的重排，它是由一摞硬币经过一个整摞翻面后得到的。因为同样的翻面过程在每一个整摞翻面之后重复进行，所以经过一连串整摞翻面之后得到的重排可以由取第一次整摞翻面后的硬币位置的置换的方幂而得到。我们可以记录硬币的移动如表 6 所示。

表 6

1	1	2	3	4
2	2	1	1	2
3	3	3	2	1
4	4	4	4	3

在第1次整摞翻面后（表6中的最后一列），硬币1在位置3，硬币2仍在位置2，等等。其中的位置是从摞顶往下数的。表示在第1次整摞翻面后硬币位置变化的置换可记成

$$\begin{pmatrix}1&2&3&4\\3&2&4&1\end{pmatrix}。$$

可以发现，用这种置换的逆置换（即将两行互换）将更方便，即

$$\begin{pmatrix}1&2&3&4\\3&2&4&1\end{pmatrix}^{-1}=\begin{pmatrix}3&2&4&1\\1&2&3&4\end{pmatrix}$$

$$=\begin{pmatrix}1&2&3&4\\4&2&1&3\end{pmatrix}。$$

试比较一下上式最右端一项的第2行与表6中的最后一列。定义φ_M是从一个整摞翻面到下一个整摞翻面硬币位置1,2,…,M的置换的逆置换，因此，移到从摞顶数起第n个位置的硬币就是开始时在位置$n\varphi_M$上的那枚硬币。如果具体进行翻面，很快就能看到φ_M中的模式。例如

$$\varphi_8=\begin{pmatrix}1&2&3&4&5&6&7&8\\8&6&4&2&1&3&5&7\end{pmatrix},$$

$$\varphi_9=\begin{pmatrix}1&2&3&4&5&6&7&8&9\\9&7&5&3&1&2&4&6&8\end{pmatrix}.$$

对这些φ_M观察到的模式，猜测有下面的引理。

引理4

$$n\varphi_M=\begin{cases}M+2-2n, & n<\dfrac{M}{2}+1,\\[2ex] 2n-M-1, & n\geqslant\dfrac{M}{2}+1.\end{cases}$$

这条引理的证明留给读者；它可以对M用归纳法而证得。这里，我们举例说明已知 φ_8后如何求 φ_9。在计算 φ_9时，在涉及到第 9 枚硬币之前，翻面时摞中硬币位置的变化与$M=8$时的变化相同。因而对$M=9$时一次整摞翻面可表示如下：

1	8	9
2	6	7
3	4	5
4	2	3
5	1	1
6	3	2
7	5	4
8	7	6
9	9	8

其中的倒数第 2 列由 φ_8得知，最后一列就给出了 φ_9。

同样的推理能用于发现：从一次整摞翻面到下一次时，最终哪些硬币的面被翻过来了。这就是

引理5 移到第 n 个位置的硬币被翻面的充分必要条件是$n<M/2+1$。

详细证明留给读者。

如果对 $2n$ 张扑克牌进行洗牌，按下面这种方式：将第 2 张牌放在第 1 张的上面，将第 3 张放在这 2 张的下面，而第 4 张放在这 3 张的上面，这样做下去。与洗这堆牌相联系的

置换恰好是 φ_{2n}. 这一类洗牌问题的理论出现在1773年([3])。Rouse Ball([4])提到讨论这一洗牌问题的 8 个人。

每一个置换可以写成若干个不相交的循环的乘积。例如

$$\varphi_8 = (1\ 8\ 7\ 5)(2\ 6\ 3\ 4),$$

$$\varphi_9 = (1\ 9\ 8\ 6\ 2\ 7\ 4\ 3\ 5).$$

任意置换的阶是这若干个不相交的循环的阶的最小公倍数。与我们讨论的硬币问题有关的 φ_M 的阶见表 7。可以看到 φ_M

表 7

M	φ_M的循环的阶	k(φ_M的阶)	Mk	M个硬币的翻面次数
8	4,4	4	32	31
9	9	9	81	80
10	6,3,1	6	60	60
11	11	11	121	121
12	10,2	10	120	119
13	9,3,1	9	117	116
14	14	14	196	195
15	5,5,5	5	75	75

的阶等于它的某一个循环的阶。这就是引理 6 的结论。

引理6 置换 φ_M 的阶等于包含数 M 的那个循环的阶。

证明 由引理 4，置换 φ_M 可以写成 $n\varphi_M = (-1)^{\varepsilon}(2n - M - 1) + \varepsilon$，其中 $\varepsilon = 0$ 或 1，从而有 $2(n\varphi_M) = (-1)^{\varepsilon}2(2n-1) - 2(-1)^{\varepsilon}M + 2\varepsilon$。由此得 $2(n\varphi_M) - 1 = (-1)^{\varepsilon}2(2n-1) - (-1)^{\varepsilon}(2M+1)$，因而有

$$2(n\varphi_M) - 1 \equiv (-1)^{\varepsilon}2(2n-1) \pmod{2M+1}. \quad (1)$$

定义

$$\chi_0(n)=2n-1,$$
$$\chi_1(n)=2(n\varphi_M)-1,$$
$$\chi_2(n)=2(n\varphi_M^2)-1,$$
$$\cdots\cdots\cdots\cdots$$
$$\chi_k(n)=2(n\varphi_M^k)-1.$$

由(1)的同余性知，存在 $\varepsilon_i=0$ 或 1 使得

$$\chi_1(n)\equiv(-1)^{\varepsilon_1}2\chi_0(n)(\bmod(2M+1)),$$
$$\chi_2(n)\equiv(-1)^{\varepsilon_2}2\chi_1(n)(\bmod(2M+1)),$$
$$\cdots\cdots\cdots\cdots$$
$$\chi_k(n)\equiv(-1)^{\varepsilon_k}2\chi_{k-1}(n)(\bmod(2M+1)).$$

由同余式性质依次得

$$\chi_2(n)\equiv(-1)^{\varepsilon_1+\varepsilon_2}2^2\chi_0(n)(\bmod(2M+1)),$$
$$\chi_3(n)\equiv(-1)^{\varepsilon_1+\varepsilon_2+\varepsilon_3}2^3\chi_0(n)(\bmod(2M+1)),$$
$$\cdots\cdots\cdots\cdots$$
$$\chi_k(n)\equiv(-1)^{\sum\varepsilon_i}2^k\chi_0(n)(\bmod(2M+1)).$$

因此，若 $2^k\equiv\pm1(\bmod(2M+1))$，则 $\chi_K(n)\equiv\pm\chi_0(n)(\bmod(2M+1))$。其中的“$-$”号不可能出现，否则

$$\chi_k(n)\equiv2(n\varphi_M^k)-1\equiv-(2n-1)(\bmod(2M+1)),$$

从而 $n\varphi_M^k\equiv1-n(\bmod(2M+1))$（因为 $2M+1$ 是奇数）。但上式左边不大于 M 而右边（即 $(1-n)$ 对模 $(2M+1)$ 的最小非负剩余）大于 M。因此得，若 $2^k\equiv\pm1(\bmod(2M+1))$，则必有 $n\varphi_M^k\equiv n(\bmod(2M+1))$。另外，$1\varphi_M^k\equiv1(\bmod(2M+1))$ 当且仅当 $2^k\equiv\pm1(\bmod(2M+1))$（为什么?）。这证明了，若 k 是包含 1 的循环的阶，则它也是置换 φ_M 的阶。因为包含 1 的循环必定也包含 M（为什么?），所以引理 6 得证。

有意思的是，φ_M 的阶就是猜想 1 中的 k。这意味着一列

都是正面朝上或一列都是背面朝上仅当所有的硬币都重新回到它们原始位置或者正好是倒序时才出现。

由考虑包含M的循环的结构，我们能得到有关M及φ_M的阶之间联系的更进一步的信息。由引理4，包含M的循环可以写成最后一个元素是1，而前面的若干个元素可以按

$$M, M-1, M-1-2, M-1-2-2^2, \cdots$$

形式继续写下去直到某一个整数$<M/2+1$为止。设该循环的第l项是小于$M/2+1$的第一个整数，则第$(l-1)$项是

$$M-(1+2+\cdots+2^{l-3})=M-2^{l-2}+1 \geqslant M/2+1,$$

第l项是

$$M-(1+2+\cdots+2^{l-2})=M-2^{l-1}+1<M/2+1.$$

由上述两个不等式得$2^{l-1} \leqslant M \leqslant 2^l-1$。如果$M=2^{l-1}$则第$l$项就是1，从而$\varphi_M$的阶是$l$，在所说的$M$的二进制范围中它是最小可能的阶数，从而猜想3证实。注意，猜想3中的其余部分已由猜想1的讨论而证得。

最后还需证明引理1。用n_k表示二进小数$p_M(n)$中的第k个数字。我们用归纳法证明，对所有k，n_k等于$X=(2M+2-2n)/(2M+1)$的二进展开式中的第k个数字。

对$k=1$，由引理5得$n_k=0$当且仅当$n \geqslant M/2+1$，或等价地，$2M+2-n<M$。这对X的二进展开中的第一个数字是零是同样的条件。假定结论对k成立。因为移到第n个位置上的硬币是$n\varphi_M$而且它被翻面的充要条件是$n<M/2+1$，所以

$$n_{k+1}=\begin{cases}(n\varphi_M)_k, & n \geqslant M/2+1,\\ 1-(n\varphi_M)_k, & n<M/2+1.\end{cases}$$

如果$n \geqslant M/2+1$，则$(n\varphi_M)_k=0$当且仅当

$$2^{k-1}(2M+2-2(2n-M-1))\leqslant M(\bmod(2M+1))$$

成立，即 $2^k(2M+2-2n)\leqslant M\ (\bmod(2M+1))$，它是$X$的二进展开式中第$(k+1)$个数字是0的充要条件。

如果 $n<M/2+1$，则 $1-(n\varphi_M)_k=1$ 当且仅当

$$2^{k-1}(2M+2-2(M+2-2n))\leqslant M(\bmod(2M+1))$$

成立，即 $2^k(2M+2-2n)>M(\bmod(2M+1))$，它是$X$的二进展开式中第$(k+1)$个数字是1的充要条件。

这就证明了，n_{k+1}是X的二进展开式中的第$(k+1)$个数字，从而由归纳法，引理1得证。

参 考 文 献

[1] Graham Birtwistle, Ole-Johan Dahl, Bjorn Myhrhaug, Kristen Nygaard, Simula Begin, Auerback, Philadelphia, PA, 1973.

[2] Trygve Nagell, Introduction to Number Theory, Chelsea, New York, 1964.

[3] Mémoires de l'Académie des Sciences, Paris, 1773.

[4] W.W. Rouse Ball, Mathematical Recreations and Essays, Macmillan, London, 1959, p.310.

（朱学贤译，潘承彪校）

圆中的蒲丰针问题[①]

M.F.NEUTS, P.PURDUE

1. 引言

蒲丰（Buffon）在他于1777年发表的题为“算术修养随笔”（Essai d'arithmétique morale）的文章中认为，几何概率应成为概率论的新分支，并提出了当今人所共知的蒲丰针问题（Buffon's needle problem）：

“地板上刻有一簇间距为$2a$的平行线。将一长度为$2c$的棒随机地投向地板，这里$2c<2a$，求此棒与平行线簇相交的概率。”

蒲丰给出了这一问题的正确答案，所求概率为$2c/(\pi a)$。分析他的解法，我们认为其基本点为：

以x表示从棒的中心到其中一条直线的距离，θ表示棒与平行线方向之间的夹角。可以假定x,θ相互独立并分别服从$(0,a)$，$(-\pi/2,\pi/2)$上的均匀分布，此即蒲丰所说的将棒“随机”地投向地板的含义。

拉普拉斯（Laplace）在其1812年的著作《概率的理论分析》(Théorie analytiqne des Probabilités）中对蒲丰针问题作了推广：考虑平面上由两组间距分别为a及b的平行线垂直相交所构成的网格，向该网格随机地投掷一长度为l（$l<a$或b）的棒，这里，随机的意义与前面的解释相同．拉普

① BUFFON IN THE ROUND, *Mathematics Magazine*, Mar.-Apr. (1971), 81—89.

拉斯给出该棒与网格中某直线相交的概率为$[2l(a+b)-l^2]/(\pi ab)$。很明显，若令a或b为无穷，就是原来的蒲丰针问题。有关这方面的进一步的研究参阅[1],[2],[3]。

本文中将讨论蒲丰针问题的一个新的推广。我们所考虑的不是平行线簇，而是一个半径为R的圆。向该圆随机投掷一根长为$2d$的针并使针的中点落在圆周内（包括圆周上）。这里，仍然需要解释所谓“随机”的含义。可以有两种不同的解释。称第一种解释为情形A，即在投掷之前首先指定一个向量径（即与某半径重合的向量——译者注），而投针的中点落在该向量径上。令U表圆心到投针中点的距离，则U服从$(0,R)$上的均匀分布，投针与向量径之间的夹角θ服从$[0,\pi]$上的均匀分布并且U与θ相互独立。第二种解释称为情形B：即针的中点落入圆内任一子区域的概率等于这部分的面积与整个圆的面积之比。对θ的假定同情形A。

确切地说，我们对随机的两种解释分别为：

A．令U表圆心到M（针的中点）的距离，则U的密度函数为

$$f(u)=\begin{cases}1/R, & 0\leqslant u\leqslant R,\\ 0, & \text{其它}.\end{cases}$$

B．U的分布密度为

$$g(u)=\begin{cases}2u/R^2, & 0\leqslant u\leqslant R,\\ 0, & \text{其它}.\end{cases}$$

以上两种情形均设θ服从$[0,\pi]$上的均匀分布并与U互相独立。

本文研究针与圆周的交点数的概率分布。情形A时此概

率可以由第一类及第二类椭圆积分表出，而在情形B时仅用初等函数便可表出其概率。附录1中给出了当d/R固定时上述交点数的概率值。

2. 交点概率的求法

定义随机变量z为针与圆周的交点数。当$d>2R$时显然有$P(z=2)=1$，因此只需考虑$0<d\leqslant 2R$的情况。我们分两种情形讨论：(a) $R\leqslant d\leqslant 2R$，(b) $0<d\leqslant R$。首先讨论情形(a)。

(a) $R\leqslant d\leqslant 2R$。此时又可分为两种情况。令$p_i(u)$为条件概率，即$p_i(u)=P(z=i|U=u)$，$i=0,1,2$，则依从O到M的距离u可将情形(a)分为：

(i) $0<u\leqslant d-R$。此时交点数必为2。即

$$p_0(u)=p_1(u)=0,$$

$$p_2(u)=1.$$

(ii) $d-R<u\leqslant R$。此时有（见图1）：

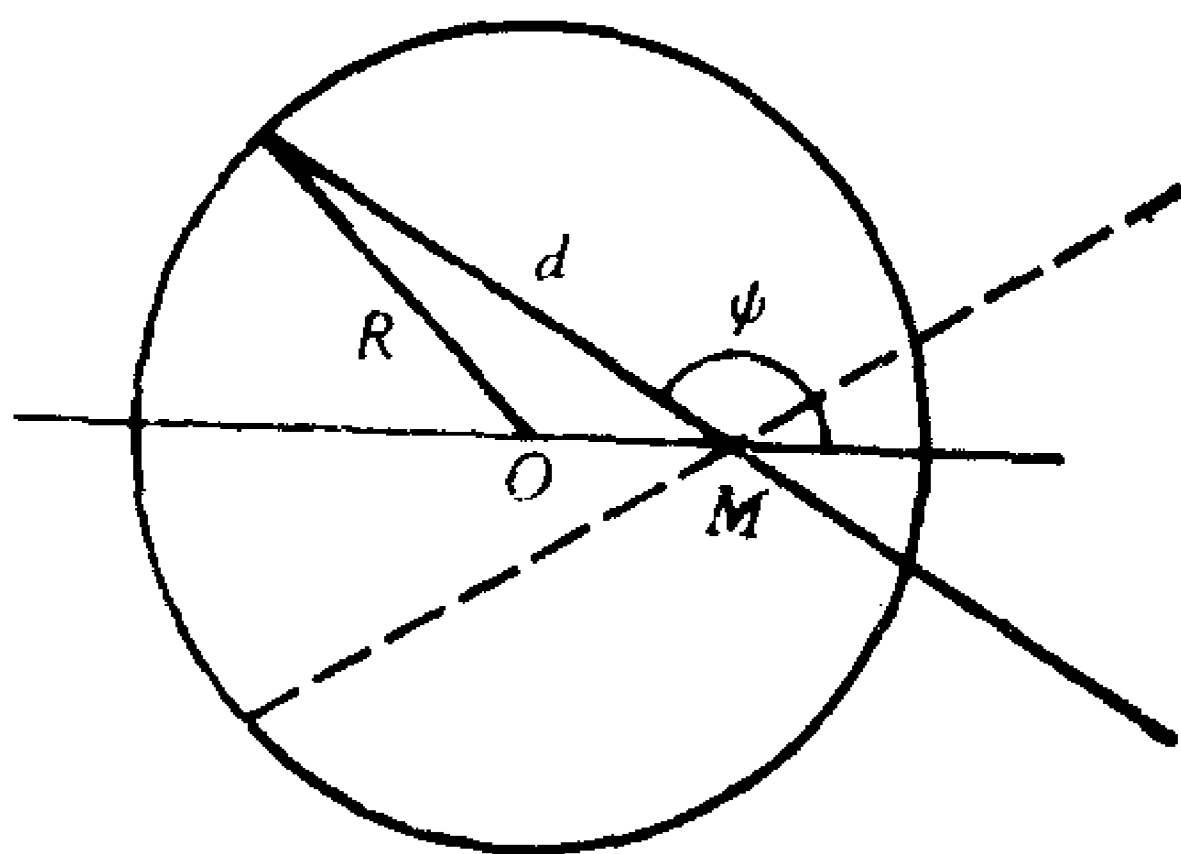


图 1

$$p_0(u)=0,$$

$$p_1(u)=2-2\psi/\pi,$$

$$p_2(u)=2\psi/\pi-1,$$

$$R^2=u^2+d^2-2ud\cos(\pi-\psi).$$

由最后一式知ψ可由u,d及R表示，因此有

$$p_0(u)=0,$$

$$p_1(u)=2-2/\pi\ \mathrm{arc}\cos\left(\frac{R^2-d^2-u^2}{2du}\right),$$

$$p_2(u) = 2/\pi \arccos\left(\frac{R^2 - d^2 - u^2}{2du}\right) - 1.$$

以 $P_A(z=i)$，$P_B(z=i)$分别记在情形 A 及 B 时交点数等于 i 的概率。则当 $R \leqslant d \leqslant 2R$ 时有

$$P_A(z=0) = 0,$$

$$P_A(z=1) = \int_{d-R}^{R} \left[2 - \frac{2}{\pi} \arccos\left(\frac{R^2 - d^2 - u^2}{2ud}\right)\right] \frac{du}{R},$$

$$P_A(z=2) = \int_0^{d-R} \frac{du}{R}$$

$$+ \int_{d-R}^{R} \left[\frac{2}{\pi} \arccos\left(\frac{R^2 - d^2 - u^2}{2ud}\right) - 1\right] \frac{du}{R},$$

$$P_B(z=0) = 0,$$

$$P_B(z=1) = \int_{d-R}^{R} \left[2 - \frac{2}{\pi} \arccos\left(\frac{R^2 - d^2 - u^2}{2ud}\right)\right] \frac{2udu}{R^2},$$

$$P_B(z=2) = \int_0^{d-R} \frac{2udu}{R^2}$$

$$+ \int_{d-R}^{R} \left[\frac{2}{\pi} \arccos\left(\frac{R^2 - d^2 - u^2}{2ud}\right) - 1\right] \frac{2udu}{R^2}.$$

化简得

$$P_A(z=0) = 0,$$

$$P_A(z=1) = 4 - \frac{2d}{R}$$

$$- \frac{2}{\pi R} \int_{d-R}^{R} \arccos\left(\frac{R^2 - d^2 - u^2}{2ud}\right) du,$$

$$P_A(z=2)=\frac{2d}{R}-3+\frac{2}{\pi R}\int_{d-R}^{R}\arccos\left(\frac{R^2-d^2-u^2}{2ud}\right)du,$$

$$P_B(z=0)=0,$$

$$P_B(z=1)=\frac{2d}{R^2}(2R-d)$$

$$-\frac{4}{\pi R^2}\int_{d-R}^{R}u\arccos\left(\frac{R^2-d^2-u^2}{2ud}\right)du,$$

$$P_B(z=2)=1-\frac{2d}{R^2}(2R-d)$$

$$+\frac{4}{\pi R^2}\int_{d-R}^{R}u\arccos\left(\frac{R^2-d^2-u^2}{2ud}\right)du.$$

上述积分的计算在第三节中讨论。

(b) $0<d\leqslant R$。依O到M之间的距离又可分解成如下三种情况：

(i) $0<u<R-d$。此时没有交点，故

$$p_0(u)=1,\quad p_1(u)=p_2(u)=0.$$

(ii) $R-d\leqslant u<\sqrt{R^2-d^2}$。此时最多只有一个 交点。由图 2 知

$$p_0(u)=1-2\psi/\pi,\quad p_1(u)=2\psi/\pi,\quad p_2(u)=0,$$

$$\psi=\arccos\left(\frac{R^2-d^2-u^2}{2ud}\right).$$

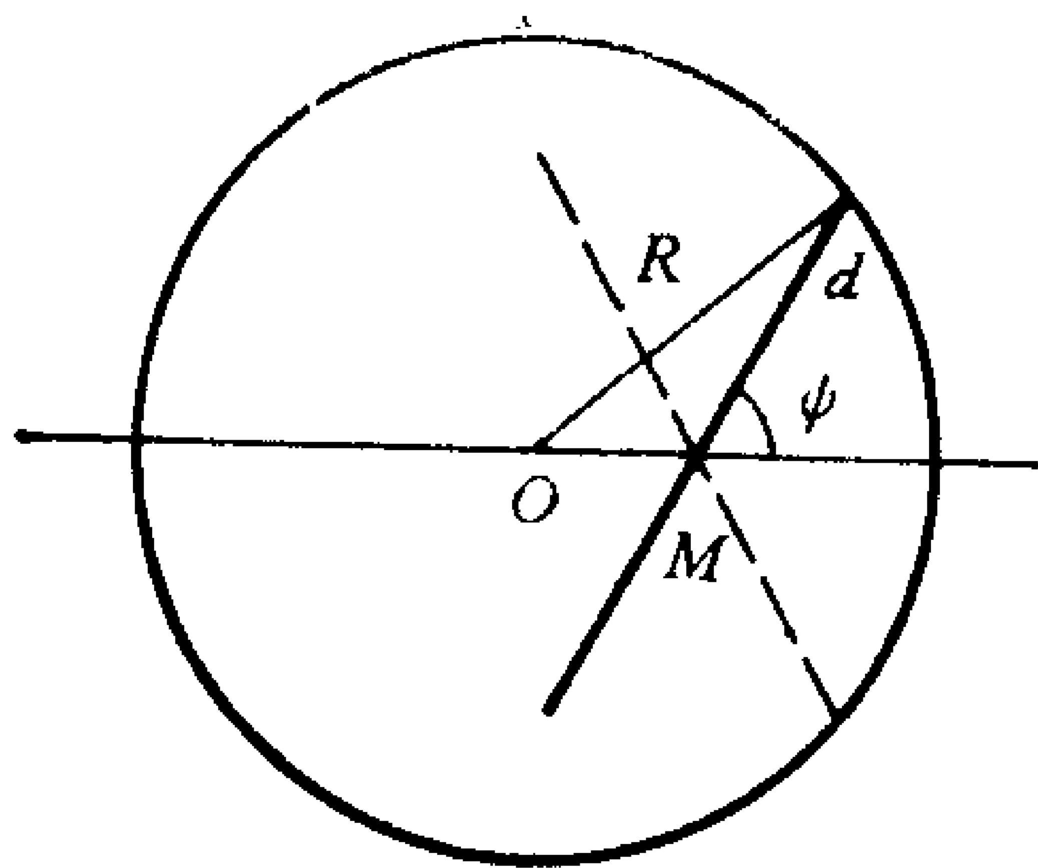


图 2

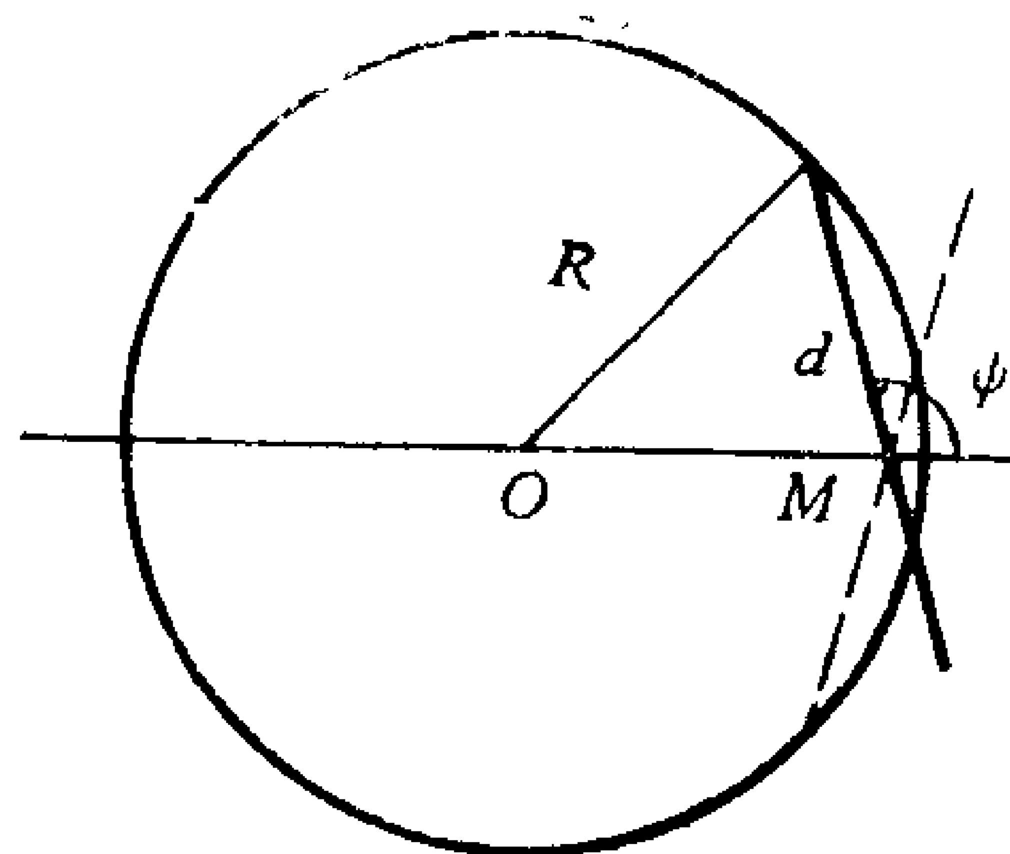


图 3

(iii) $\sqrt{R^2-d^2}\leqslant u\leqslant R$。此时至少有一个交点，并依图 3 得到

$$p_0(u)=0,\quad p_1(u)=2-2\psi/\pi,\quad p_2(u)=2\psi/\pi-1.$$

类似于情形(a)，我们有

$$P_A(z=0)=\frac{\sqrt{R^2-d^2}}{R}-\frac{2}{\pi R}\int_{R-d}^{(R^2-d^2)^{1/2}}\arccos\left(\frac{R^2-d^2-u^2}{2ud}\right)du,$$

$$P_A(z=1)=2-\frac{2\sqrt{R^2-d^2}}{R}+\frac{2}{\pi R}\int_{R-d}^{(R^2-d^2)^{1/2}}\arccos\left(\frac{R^2-d^2-u^2}{2ud}\right)du-\frac{2}{\pi R}\int_{(R^2-d^2)^{1/2}}^{R}\arccos\left(\frac{R^2-d^2-u^2}{2ud}\right)du,$$

$$P_A(z=2)=\frac{\sqrt{R^2-d^2}}{R}-1+\frac{2}{\pi R}\int_{(R^2-d^2)^{1/2}}^{R}\arccos\left(\frac{R^2-d^2-u^2}{2ud}\right)du.$$

$$P_B(z=0)=\frac{R^2-d^2}{R^2}-\frac{4}{\pi R^2}\int_{R-d}^{(R^2-d^2)^{1/2}}u\arccos\left(\frac{R^2-d^2-u^2}{2ud}\right)du,$$

$$P_B(z=1)=2-\frac{2}{R^2}(R^2-d^2)+\frac{4}{\pi R^2}\int_{R-d}^{(R^2-d^2)^{1/2}}u\arccos\left(\frac{R^2-d^2-u^2}{2ud}\right)du-\frac{4}{\pi R^2}\int_{(R^2-d^2)^{1/2}}^{R}u\arccos\left(\frac{R^2-d^2-u^2}{2ud}\right)du,$$

$$P_B(z=2)=\frac{R^2-d^2}{R^2}-1+\frac{4}{\pi R^2}\int_{(R^2-d^2)^{1/2}}^{R}u\arccos\left(\frac{R^2-d^2-u^2}{2ud}\right)du.$$

3．积分的计算

令

$$I=\int_a^b \arccos\left(\frac{R^2-d^2-u^2}{2ud}\right)du.$$

通过分部积分及下列变量代换

$$u=(R+d)\cos\phi,$$

$$(R+d)\sin\phi=2\sqrt{Rd}\sin\theta$$

得到

$$I=u\arccos\left(\frac{R^2-d^2-u^2}{2ud}\right)\Bigg|_a^b$$

$$-\{(d-R)[F(b_2|k)-F(a_2|k)]$$

$$-(d+R)[E(b_2|k)-E(a_2|k)]\},$$

其中

$$F(\eta|k)=\int_0^{\eta}(1-k^2\sin\ \theta)^{-\frac{1}{2}}d\theta,$$

$$E(\eta|k)=\int_0^{\eta}\sqrt{1-k^2\sin^2\theta}\,d\theta,$$

$$k^2=4Rd/(R+d)^2,$$

$a_2,b_2=$ 上述积分关于 θ 的积分限。

可见，$F(\eta|k)$ 为第一类椭圆积分，而 $E(\eta|k)$ 为第二类椭圆积分。令

$$I'=\int_a^b u\ \mathrm{arc}\cos\left(\frac{R^2-d^2-u^2}{2ud}\right)du。$$

类似于 I 的计算过程，可得

$$I'=\frac{u^2}{2}\mathrm{arc}\cos\left(\frac{R^2-d^2-u^2}{2ud}\right)\Big|_a^b$$

$$+R^2(b_2-a_2)+Rd\sin\theta\cos\theta\ \Big|_{a_2}^{b_2}。$$

4．主要结果

利用第 3 节的结果可得到关于 $P_A(z=i)$，$P_B(z=i)$ 的表达式。令

$$\gamma=d/R,\qquad \theta_1=\pi/2,$$

$$\theta_2=\mathrm{arc}\sin\sqrt{\frac{1+\gamma}{2}},\qquad \theta_3=\mathrm{arc}\sin\sqrt{\frac{2+\gamma}{4}}。$$

关于情形 A 的结果：

(i) $0<\gamma\leqslant1$。

$$P_A(z=0)=\frac{2}{\pi}\{(1-\gamma)[F(\theta_1|k)-F(\theta_2|k)]$$

$$+(1+\gamma)[E(\theta_1|k)-E(\theta_2|k)]\},$$

$$P_A(z=1)=\frac{2}{\pi}\text{arc cos}\,\frac{\gamma}{2}$$

$$-\frac{2}{\pi}\{(1-\gamma)[F(\theta_1|k)-F(\theta_2|k)]$$

$$+(1+\gamma)[E(\theta_1|k)-E(\theta_2|k)]\}$$

$$\frac{2}{\pi}\{(1-\gamma)[F(\theta_2|k)-F(\theta_3|k)]$$

$$+(1+\gamma)[E(\theta_2|k)-E(\theta_3|k)]\},$$

$$P_A(z=2)=1-\frac{2}{\pi}\text{arc cos}\,\frac{\gamma}{2}$$

$$-\frac{2}{\pi}\{(1-\gamma)[F(\theta_2|k)-F(\theta_3|k)]$$

$$+(1+\gamma)[E(\theta_2|k)-E(\theta_3|k)]\}.$$

(ii) $1\leqslant\gamma\leqslant 2$.

$$P_A(z=0)=0,$$

$$P_A(z=1)=\frac{2}{\pi}\text{arc cos}\,\frac{\gamma}{2}$$

$$-\frac{2}{\pi}\{(\gamma-1)[F(\theta_1|k)-F(\theta_2|k)]$$

$$-(\gamma+1)[E(\theta_1|k)-E(\theta_3|k)]\},$$

$$P_A(z=2)=1-\frac{2}{\pi}\text{arc cos}\,\frac{\gamma}{2}$$

$$+\frac{2}{\pi}\{(\gamma-1)[F(\theta_1|k)-F(\theta_3|k)]$$

$$-(\gamma+1)[E(\theta_1|k)-E(\theta_3|k)]\}.$$

因 k 可由 γ 表示，即 $k^2=4\gamma/(1+\gamma)^2$，故上述各表达式均可看成关于 γ 的单变量函数来计算。

关于情形 B 的结果：

(i) $0<\gamma\leqslant1$。

$$P_B(z=0)=2-\frac{2\gamma}{\pi}\sqrt{1-\gamma^2}-\frac{4}{\pi}\arcsin\sqrt{\frac{1+\gamma}{2}},$$

$$P_B(z=1)=\frac{4}{\pi}\sqrt{1-\gamma^2}-\frac{\gamma}{4}\sqrt{4-\gamma^2}-2$$

$$+\frac{2}{\pi}\arccos\frac{\gamma}{2}+\frac{8}{\pi}\arcsin\sqrt{\frac{1+\gamma}{2}}$$

$$-\frac{4}{\pi}\arcsin\sqrt{\frac{2+\gamma}{4}},$$

$$P_B(z=2)=1-\frac{2\gamma}{\pi}\left(\frac{1}{2}\sqrt{4-\gamma^2}-\sqrt{1-\gamma^2}\right)$$

$$-\frac{2}{\pi}\arccos\frac{\gamma}{2}+\frac{4}{\pi}\arcsin\sqrt{\frac{2+\gamma}{4}}$$

$$-\frac{4}{\pi}\arcsin\sqrt{\frac{1+\gamma}{2}}.$$

(ii) $1\leqslant\gamma\leqslant2$。

$$P_B(z=0)=0,$$

$$P_B(z=1)=2-\frac{\gamma}{\pi}\sqrt{4-\gamma^2}+\frac{2}{\pi}\arccos\frac{\gamma}{2}$$

$$-\frac{4}{\pi}\arcsin\sqrt{\frac{2+\gamma}{4}},$$

$$P_B(z=2)=\frac{\gamma}{\pi}\sqrt{4-\gamma^2}-1+\frac{4}{\pi}\arcsin\sqrt{\frac{2+\gamma}{4}}$$

$$-\frac{2}{\pi}\arccos\frac{\gamma}{2}.$$

用计算器很容易计算上述结果．附录 1 中给出了一些具体的数值，其中 γ 的取值介于 0 到 2 之间。

5．附录

图 4 给出了当$\gamma\in[0,2]$时各概率之间的关系，其中情形 A 以实线表示，而情形 B 则用虚线表示．表 1 中列出了情形 A 时的计算结果，而情形 B 时的计算结果则在表 2 中给出。

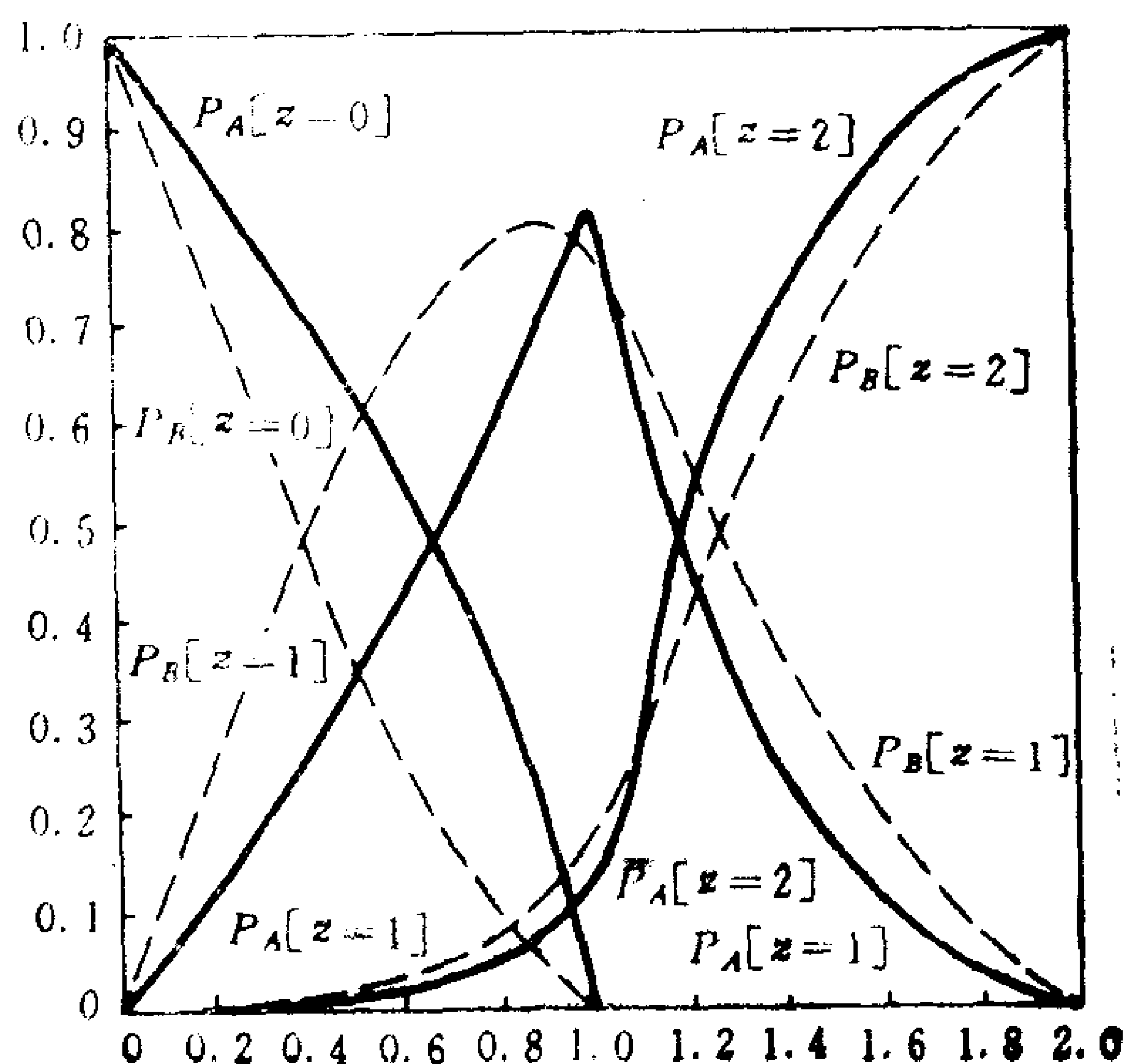


图 4

表1 关于情形A的计算结果

γ	$P_A[z=0]$	$P_A[z=1]$	$P_A[z=2]$
.2	.86	.14	.00
.4	.70	.29	.01
.6	.52	.46	.02
.8	.30	.64	.06
1.0	.00	.84	.16
1.2	.00	.45	.55
1.4	.00	.22	.78
1.6	.00	.13	.87
1.8	.00	.04	.96

表2 关于情形B的计算结果

γ	$P_B[z=0]$	$P_B[z=1]$	$P_B[z=2]$
.2	.75	.25	.00
.4	.51	.48	.01
.6	.28	.68	.04
.8	.11	.80	.09
1.0	.00	.78	.22
1.2	.00	.57	.43
1.4	.00	.38	.62
1.6	.00	.21	.79
1.8	.00	.07	.93

参 考 文 献

[1] C.B. Boyer, A History of Mathematics, Wiley, New York, 1968.

[2] M.W. Crofton, Probability, Encylopedia Britannica,

9th ed ,(1885).

[3] J.J. Sylvester, On Buffon's problem of the needle, *Acta Math.*, 14 (1891) , 185--205.

[4] A.L. Clarke, Buffon's needle problem, *Canad.J. Res.*, 9 (1933) , 402 and 11 (1934) , 658.

[5] N. T. Gridgeman, Geometric probability and the number π, *Scripta Math.*, 25 (1960) , 183—195.

[6] B. C. Kahan, A practical demonstration of a needle experiment to give a number of concurrent estimates for π, *J. Roy. Statist. Soc. Ser.* A, 124(1961),227—239.

[7] M. S. Klamkin, On the Uniqueness of the distribution function for the Buffon needle problem, *Amer. Math. Monthly*, 60 (1953) , 677—680.

[8] ——, On Barbier's solution of the Buffon needle problem, this Magazine, 28 (1955) , 135—138.

[9] L. Mantel,An extension of the Buffon needle problem, Ann. Math. Statist., 22 (1951),314—315,also *Ann.Math. Statist.*, 24 (1953) 674--677.

[10] J. F. Ramaley, Buffon's noodle problem, *Amer. Math. Monthly*, 76 (1969) , 916—918.

（范永亮译，朱学贤校）

国际象棋中的“n王问题”[①]

M. Abramson, W. Moser

设$A=(a_{ij})$是一个$n\times m$矩阵，有nm个不同的元素。从中选取k个，使两两不同行，而且在相邻行的两个元素必在同一列，记这样的选取方法共有$g_{n,k}(m)$种。本文的主要目的是用初等方法求出$g_{n,k}(m)$，并利用$g_{n-1,k}(2)(k=0,1,2,\cdots)$去解“$n$王问题”：在一个$n\times n$的棋盘中放入$n$个王，使每行、每列都只有一个，并且两两不能相互攻击，问有多少种放置方法？后面这个问题的本身也很有趣。

我们称k个选自集合$\{1,2,\cdots,n\}$的整数

$$x_1<x_2<\cdots<x_k \tag{1}$$

为n的一个k-选择。其中由相继的整数组成的最长的一段称为k-选择的一个部分。例如

$$1,3,4,5,8,9 \tag{2}$$

是10的一个6-选择，它共有三个部分，为(1),(3,4,5)及(8,9)，长度分别是1,3及2。

恰有r个部分的n的k-选择个数是

$$g_r(n,k)=\binom{k-1}{r-1}\binom{n-k+1}{r}. \tag{3}$$

为证明(3)式，我们用“·”表示一个数未被选取，“○”

① Combination, Successions and the n-KINGS Problem, *Mathematics Magazine*, Nov.—Dec.(1966), p.269—273.

表示一个数被选取。这样，n 的恰有 r 部分的 k- 选择就同由 k 个○和 $n-k$ 个・沿直线排列组成的恰有 r 段○的有序排列（从左到右逐渐增大）有一个一一对应。因而只需确定这种排列的个数。$n-k$ 个・之间至多有 $n-k+1$ 个空段（包括第一个・以前的空和最后一个・以后的空），从中挑出 r 个空的方法数是 $\binom{n-k+1}{r}$。然后，对每种挑法，再将 k 个○插入这 r 个空，使之都填满，共 $\binom{k-1}{r-1}$ 种放法。从而得到(3)式。

$r=k$ 时，得到 Kaplansky 引理([1])：n 的无相连数的 k-选择个数是

$$g_k(n,k)=\binom{n-k+1}{k}. \tag{4}$$

在 k- 选择(1)中，数对 x_i, x_{i+1} 称为是一个后继，如果 $x_{i+1}=x_i+1$。(1)的一个长度为 a 的部分中有 $a-1$ 个后继。因此若(1)有 r 部分，长度分别是 $a_1, a_2, \cdots, a_r$，则共有 $s=a_1+a_2+\cdots+a_r-r=k-r$ 个后继。将 $r=k-s$ 代入 (3) 式，得到 Riordan 定理([4])：n 的恰有 s 个后继的 k- 选择个数是

$$f_s(n,k)=g_{k-s}(n,k)=\binom{k-1}{s}\binom{n-k+1}{k-s}.$$

现在我们对依一个圆周排列的整数考虑类似的结果，这时，n 和 1 是相邻的两个数。对这样排列的整数，n 的恰有 r 部分的 k- 选择个数是

$$h_r(n,k)=\frac{n}{n-k}\binom{n-k}{r}\binom{k-1}{r-1}. \tag{5}$$

这是因为，n的有r部分的k-选择同由k个○和$n-k$个·组成的恰有r段○并标记某个○或·为1的圆形排列（顺时针方向增大）一一对应。我们来确定这种排列的个数。圆上的$n-k$个·恰好有$n-k$个空，从中选取r个空的方法数是$\frac{1}{n-k}\binom{n-k}{r}$（因为$\binom{n-k}{r}$种·的排法可以分成$n-k$个组，每两组之间只差一个旋转）.现在将$k$个○放入这$r$个空中，使之都填满，共$\binom{k-1}{r-1}$种放法。再对每一个由○和·组成的圆形排列，依次标记每一个○或·为1，共n种标法。即得(5)式。

$r=k$时，得到Kaplansky引理[1]：n的圆周排列无相连数的圆型k-选择个数是

$$h_k(n,k)=\frac{n}{n-k}\binom{n-k}{k}. \tag{6}$$

将$r=k-s$代入(5)式得，n的有s个后继的圆型k-选择个数为

$$h_{k-s}(n,k)=\frac{n}{n-k}\binom{n-k}{k-s}\binom{k-1}{s}.$$

下面来建立一个递推公式

$$g_r(n,k)=g_r(n-1,k)+g_{r-1}(n-2,k-1)$$
$$+g_r(n-1,k-1)-g_r(n-2,k-1). \tag{7}$$

假设n的恰有r部分的k-选择：

(i) 不含n，则也是$n-1$的恰有r部分的k-选择，共

$g_r(n-1,k)$个；

(ii) 含 n 但不含 $n-1$，删去 n 后是 $n-2$ 的恰有 $r-1$ 部分的 $(k-1)$-选择，共 $g_{r-1}(n-2,k-1)$ 个；

(iii) 含 n 且含 $n-1$，删去 n 后是 $n-1$ 的且包含 $n-1$ 的有 r 部分 $(k-1)$-选择，共 $g_r(n-1,k-1)-g_r(n-2,k-1)$ 个。因为 $g_r(n-1,k-1)$ 个 $n-1$ 的有 r 部分的 $(k-1)$-选择中有 $g_r(n-2,k-1)$ 个不含 $n-1$.

综上即得(7)式。

如果用 Riordan 的记号 $f_s(n,k)$，(7)式是[4]中的引理。

递推公式(7)等价于

$$\begin{cases} g_r(n,k)=g_r(n-1,k)+\sum_{j=1}^{k-1} g_{r-1}(n-j-1,k-j), 2\leqslant r\leqslant k, \\ g_1(n,k)=g_1(n-1,k)+1. \end{cases} \tag{8}$$

当然 $g_1(n,k)=n-k+1$。利用数学归纳法可以由(8)式得到(3)式。

我们现在来确定文章开始时提到的 $g_{n,k}(m)$。（矩阵 A 的行数）n 的每一个恰有 r 部分的 k-选择都产生了 m^r 个满足文章开始时所提条件的选取方法，因此由(3)得

$$g_{n,k}(m)=\sum_{r=1}^{k} m^r g_r(n,k)=\sum_{r=1}^{k} m^r \binom{k-1}{r-1}\binom{n-k+1}{r}. \tag{9}$$

可以看到 $g_{n,k}(m)$ 正是下式中 x^k 的系数：

$$\begin{aligned}&(1+mx+mx^2+mx^3+\cdots)^{n-k+1}\\ &\quad=[1+(m-1)x]^{n-k+1}(1-x)^{-(n-k+1)}\\ &\quad=\sum_{i=0}^{n-k+1}\binom{n-k+1}{i}(m-1)^i x^i \sum_{j=0}^{\infty}\binom{n-k+j}{j}x^j.\end{aligned}$$

令 $j=k-i$，则 x^k 的系数是

$$g_{n,k}(m)=\sum_{i=0}^{k}(m-1)^i\binom{n-i}{k-i}\binom{n-k+1}{i}. \qquad (10)$$

这个表达式也可以由(9)式得到，因为

$$\sum_{r=1}^{k}m^r\binom{k-1}{r-1}\binom{n-k+1}{r}$$

$$=\sum_{r=0}^{k}\binom{k-1}{k-r}\binom{n-k+1}{r}\sum_{s=0}^{r}\binom{r}{s}(m-1)^s$$

$$=\sum_{s=0}^{k}(m-1)^s\binom{n-k+1}{s}\sum_{r=s}^{k}\binom{n-k+1-s}{r-s}\binom{k-1}{k-r}$$

$$=\sum_{s=0}^{k}(m-1)^s\binom{n-k+1}{s}\binom{n-s}{k-s}.$$

由(7)和(9)式，我们得到递推公式

$$g_{n,k}(m)=g_{n-1,k}(m)+g_{n-1,k-1}(m)+(m-1)g_{n-2,k-1}(m). \qquad (11)$$

由于 $g_{n,k}(1)=\binom{n}{k}$，因此 $m=1$ 时（11）式为

$$\binom{n}{k}=\binom{n-1}{k}+\binom{n-1}{k-1}.$$

Kaplansky 在[2],[3]中引入的数 $A_{n,k}$ 是我们这里的

$$g_{n-1,k}(2)=\sum_{i=0}^{k}\binom{n-i+1}{k-i}\binom{n-k}{i}$$

$$=\sum_{i=0}^{k}\binom{n-k+i-1}{i}\binom{n-k}{k-i},$$

与Riordan在[5,第710页]中给出的一致。下面用它来解"n王"问题。记在 i 列上的王的所在行为 $\pi(i)$，由于n个王两两不同行且不同列，因此 $\pi(1),\pi(2),\cdots,\pi(n)$是$1,2,\cdots,n$的一个排列。又由$n$个王两两不能相互攻击，所以又有

$$|\pi(s+1)-\pi(s)|\neq 1,\quad s=1,2,\cdots,n-1.$$

我们用(i,j)来表示（任一排列中）事件"j紧随i"(在$1,2,\cdots$的任意一个排列中)。对应于n个王所允许的位置的排列一定是除去以下 $2n-2$ 个事件（排成$(n-1)\times 2$阵列）的排列：

$$\begin{array}{cc} (1,2) & (2,1) \\ (2,3) & (3,2) \\ \vdots & \vdots \\ (n-1,n) & (n,n-1) \end{array} \tag{12}$$

显然一个排列不可能同时包含(12)中同一行中的两个事件，且若同时包含(12)中两个相邻行中的两个事件，则它们必定取自一列中。因此(12)中选k个（对于排列）相容事件的方法数是 $g_{n-1,k}(2)$、并且不难看出包含(12)中k个相容情况的排列的个数是$(n-k)!$。因此，由熟知的容斥原理(Principle of Inclusion and Exclusion)得，"n王问题"的解是

$$\sum_{k=0}^{n-1}(-1)^k g_{n-1,k}(2)(n-k)!,$$

其中 $g_{n-1,0}(2)=1$。

参 考 文 献

[1] I. Kaplansky, Solution of the"problème des ménages," *Bull. Amer. Math. Soc.*, 49(1943), 784—785.

[2] ——,Symbolic solution of certain problems in permu-

tations, *Bull.Amer.Math.Soc.*, 50(1944), 906—914.

[3] ——, The asymptotic distribution of runs of consecutive elements, *Ann. Math. Statist.*, 16 (1945), 200—203.

[4] J. Riordan, Permutations without 3-sequences, *Bull. Amer. Math. Soc.*, 51 (1945), 745—748.

[5] ——, A recurrence for permutations without rising or falling successions, *Ann. Math. Statist.*, 36 (1965), 708—711.

（刘　杰译，朱学贤校）

游 历 迷 宫[1]

A.S.FRAENKEL

1。简介

迷宫是一个围起来的（通常是在地下的）一些小胡同，这些胡同按不同方式相互连通。可用一个有限的连通图来描述任意一个迷宫，图中的边代表胡同，顶点代表胡同的交叉点和死胡同的终点。今后，我们将不加区别地使用“边”和“胡同”，“顶点”和“交叉点”这样一些术语。图1是一个迷宫的例子，与其相关联的图在图2中给出，图3是图2的一个变形。

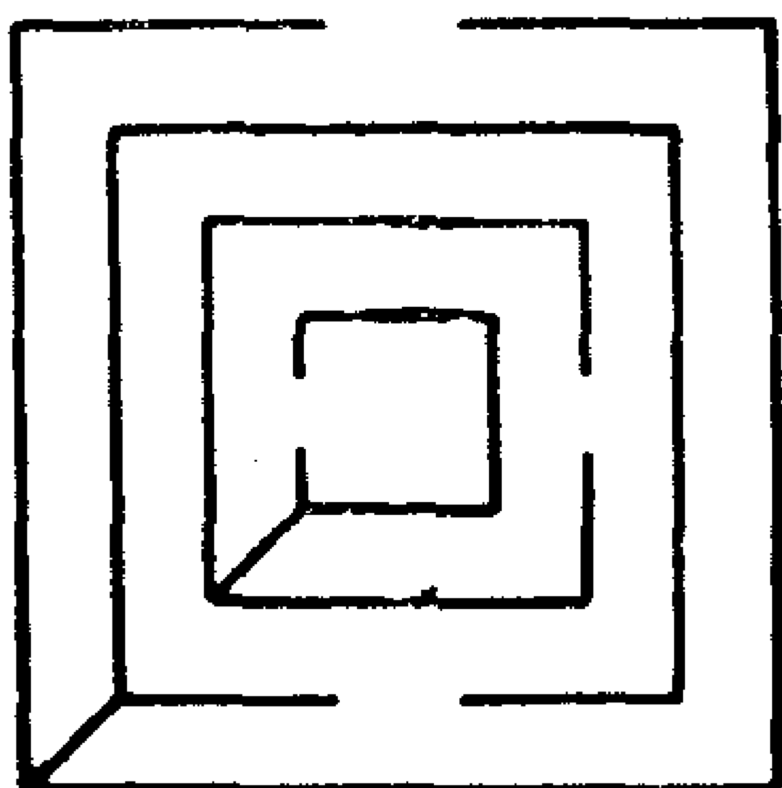


图 1

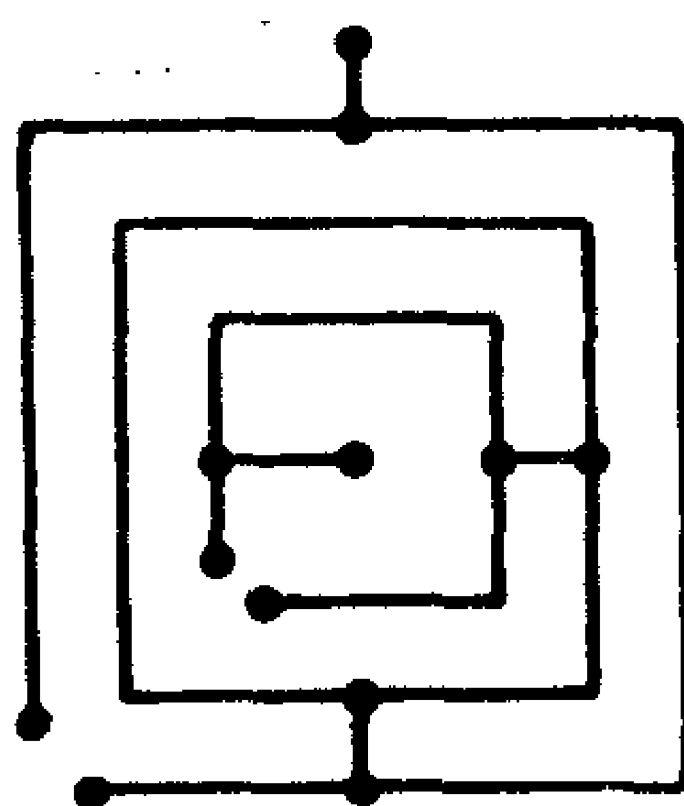


图 2

---

① Economic traversal of labyrinths, *Mathematics Magazine, May-June* (1970), 125—130.

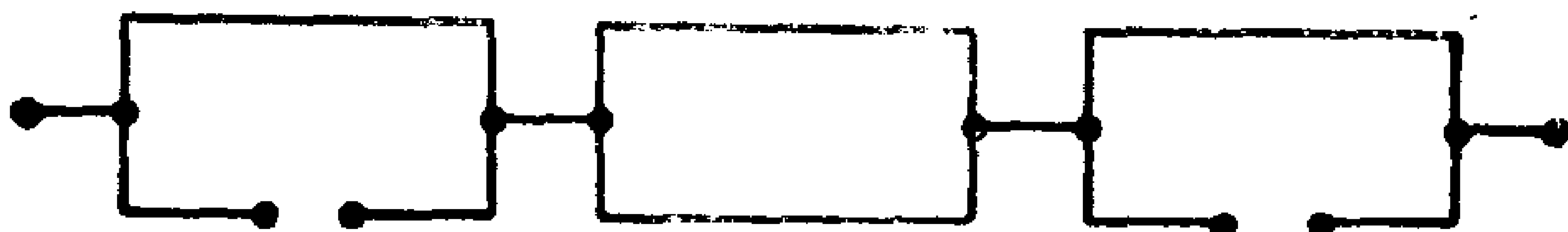


图 3

有关迷宫的一个典型问题是从一个入口走到中心，在那儿藏着财宝或一个人身牛头怪物米那多，正等着希腊神话中的英雄 Theseus 来杀死它（后来，Theseus 借助于重绕一团由 Ariadne 抓住一头的线而从迷宫脱身）；或者一个人在迷宫中迷路后，想找到一个出口。因为总是假定游历迷宫的人并不知道迷宫的构图，特别是不知道其“中心”的位置，通常要寻找的是下面更一般问题的解：从任意一个给定的顶点开始，游历图中的每条边。

存在游历了一个迷宫，而且经过了每条边的一些算法。每个这样的算法都附带着一些假设，而这些假设是关于交叉点的一些可利用的信息，即关于已走过的路径的信息。例如，Wiener算法保证每条边至少被走过两次；Trémaux和Tarry的算法保证每条边恰好走过两次，即按边的每个方向各走过一次(见[2,3])．只要代表迷宫的图是一棵树，按“沿墙走右转弯”的规则可获得同样的结果(见[1])。Trakhtenbrot 还考虑过不连通的图(迷宫)(见[4])．对一个连通图的特殊情况，他也给出了一个每条边被走过两次的算法。他认为：“能否产生比我们已有的算法更简单的算法是值得怀疑的”。

很自然地人们要寻找一个算法，该算法允许产生一个每条边至少走过一次，最多走过两次的游历。本文的目的是为了系统阐述和证明解决一般迷宫的这样一个算法的有效性，

例如，如果迷宫代表一个 Euler 图，幸运地话，可得出每条边只走过一次的游历。

由于我们的算法基于 Tarry 算法，为完整起见，我们先从证明Tarry算法开始。

2. Tarry算法

Tarry 假设，当到达任意一个交叉点 v 时，两件事情是已知的：(i) 与点 v 相关联的，且按离开 v 的方向已经走过的边的集合；(ii) 我们首次到达点 v 时所经过的边（称其为点 v 的进入边）。

在这些假设下，Tarry算法是这样的：当到达一个交叉点 v 时，继续沿着一条尚未按由 v 到v'方向走过的边 (v,v') 前进，但是，除非万不得已，不要选择进入边。

定理1 假定一个迷宫按照Tarry算法从一个起始顶点 v_0 出发进行游历，则这个游历在点v_0终止，且每条边将正好被走过两次，即按边的每个方向各一次。

证明 (i) 由于图是有限的，游历可终止。换句话说，会到达这样一个顶点 v，使每一条关联于 v 的边，按离开 v 的方向都被走过。很明显，图中的边没有按每个方向走过多于一次的。如果$v \neq v_0$，进入顶点 v 的次数比离开它的次数多 1，因而，存在一条关联于 v 的边，还没有按离开 v 的方向沿该边走过 v，这是矛盾的，因而 $v = v_0$。

(ii) 假定游历迷宫相当于在顶点

$$v_0, v_1, v_2, \cdots, v_0 \qquad (1)$$

上，按这个顺序的一个走动（除点v_0外，在这个点序列中可包含重复的点）。游历的结果，所有关联于v_0的边按离开v_0的

方向被走过一次，否则游历还可继续。因而，所有关联于v_0的边在朝向v_0的方向也被走过一次。

现在，我们证明相同的情况适用于序列(1)中的所有其他顶点。假若不然，设v_n为序列(1)中不是所有与它相关联的边都被走过的第一个顶点。因为在点v_n进入和离开的边数必须相同，特别地，存在一条与v_n关联的边，沿着这条边按离开v_n的方向未走过v_n。

假定(v_k, v_n)是v_n的进入边，那么在(1)中v_k在v_n之前出现。由于v_n的极小性，边(v_k, v_n)按从v_n到v_k的方向已被走过。但这与上面的规则相矛盾。因为规则说明仅当所有其他可能性消失时，进入边才能被用做出口。

(iii) 剩下来的只要证明(1)中包含了图中的所有顶点。设w是一个没有被走过的顶点。由于图是连通的，存在一条连接v_0和w的路

$$v_0 = w_0, w_1, \cdots, w_{m-1}, w_m = w.$$

设w_i为这条路中不包含在序列(1)中的第一个顶点，那么w_i一次也未被走过。然而，w_{i-1}是在(1)中，边(w_{i-1}, w_i)被走过。这一矛盾得出了结果。

3. 一个经济的算法

作出如下假设：到达一顶点v时，其进入边、先前已走过的关联于v的边、及其走过它们的方向均为已知。另外，游历迷宫的人有一个计数器（例如用铅笔和纸）。

设$\rho(v)$是v的价，即与v关联的边数。设v_0为初始顶点，不失一般性，可假定$\rho(v_0) = 1$。因为如果不是这种情况，我们可增加一条通向迷宫入口的短胡同，这个胡同的起

始端点的价为1。

算法：

(1) 从 v_0 开始且计数器为零。每当我们到达一个以前未走过的顶点时，计数器加1。

(2) 如果我们到达一个顶点 v，在进入该点之前，至少有一条关联于该点的边尚未走过；到达 v 后，最多仍有一条这样的边，计数器减1。

(3) 只要计数器是正的，可根据Tarry算法指导游历。但是只要可能，最好使用一条以前从未走过的边，而不用以前已经走过的边。

(4) 一旦计数器为零，沿着它们的进入边离开所有的顶点。

定理2 游历在 v_0 结束，使得每条边最少走过一次，且在边的每个方向上最多走过一次。

例1 考虑图4给出的用图表示的迷宫，其顶点已用数字0—7标号，初始顶点标号为零。根据算法可得到一个游历，它所经过的顶点序列为：

0,1,2,3,1,3,4,5,6,4,7,2,5,7,3,
7,5,2,7,4,6,6,5,4,3,2,1,0。

图 4

读者很容易验证，最后一次出现的顶点 6 代表计数器为零的点。从这时起，旅行相当于沿进入边返回。在图 4 及后面的各图中，所有进入边用粗线表示。注意，这个游历比按照 Trémaux，Tarry或Trakhtenbrot的游历仅有的节省是在顶点 6 的环上，只走过一次。

例 2　对同一个迷宫的，与算法一致的另外一个走法，可由序列

$$0,1,2,5,6,6,4,5,7,4,3,7,2,3,1,0$$

给出（见图 5）。可见，这是一个最经济的游历，因为，所有

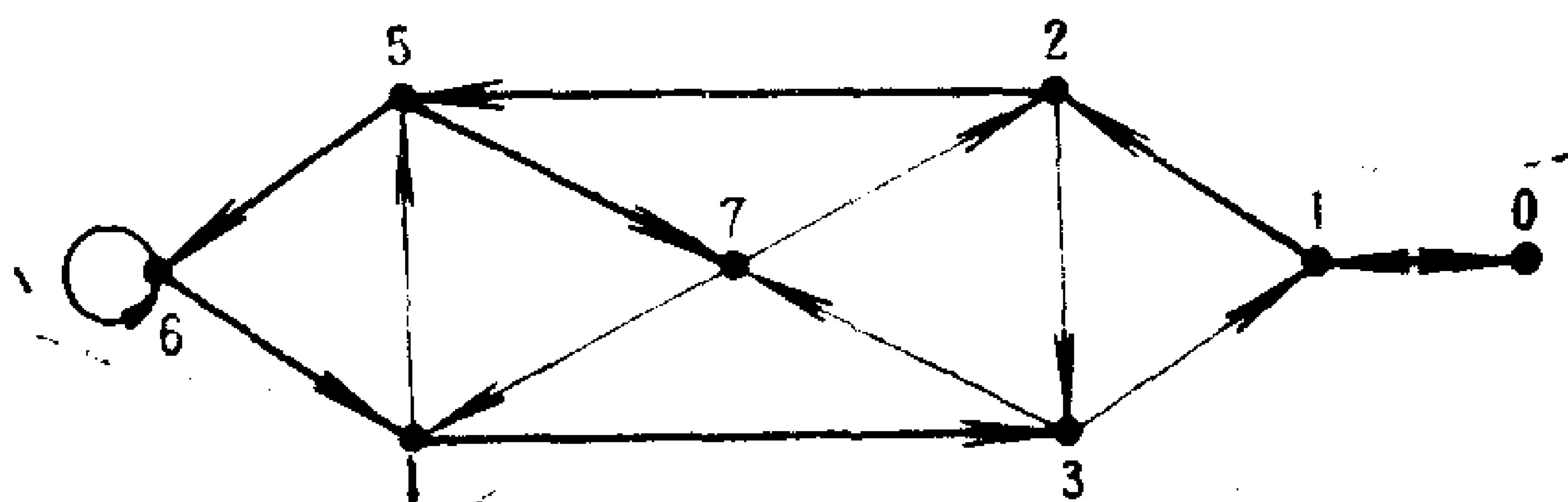


图　5

与偶价顶点相关联的边仅走过一次，只有连结顶点 0 和 1 的边（0 和 1 是奇价点）走过两次。

定理2的证明　(i) 首先我们证明计数器将最终为零。假若不然，那么仅仅是根据 Tarry 算法游历迷宫。如果迷宫的顶点为 $v_0, v_1, \cdots, v_n$，n 个单位加进了计数器，且在游历过程中又减去 n 个单位，这样计数器最终为零。

(ii) 其次，我们证明当计数器为零时，所有边至少被走过了一次。如果需要，可重新命名顶点，我们假定在走过了顶点 $v_0, v_1, \cdots, v_k$ 后，计数器首次变为零。显然，与点 v_i

$(0 \leqslant i < k)$ 关联的所有边一定已经走过。换言之，不存在这样的顶点，与其相关联的仅仅部分边被走过。假设 w 是一个一次也未走过的顶点，存在一条连接 v_0 与 w 的简单路 $v_0 = w_0, w_1, \cdots, w_{m-1}, w_m = w_0$，设 w_i 为该序列中未被走过的第一个顶点，那么 w_{i-1} 已被走过，特别地，边 (w_{i-1}, w_i) 被走过，这是一个矛盾。

(iii) 在边的每个方向上走过不会超过一次。这对计数器为正数的游历部分是明显的，因为这一部分是按 Tarry 算法游历的子集。剩下来的只要证明，如果 P 是到达那儿时计数器为零的交叉点，游历的其余部分亦符合结论。用归纳法证明。由 Tarry 算法，P 的进入边尚未按离开 P 的方向走过，而现在我们可这样做了。假定在游历过程中，我们沿点 q 的进入边 (q, r) 离开 q，在这之前，该边按 q 到 r 的方向还未走过。如果这是我们第一次访问 r 且计数器为零，Tarry 算法保证 r 的进入边尚未被用作一个出口。假设在计数器变为零后，我们已经访问过 r。那么存在一条进入边的回路

$$r = w_0, w_1, \cdots, w_{m-1}, w_m = w_0.$$

假定 w_i 为具有正计数的游历期间被进入的那些点中的第一个点，那么沿边 (w_{i-1}, w_i) 进入 w_i。这意味着 w_{i-1} 在它之前就被进入了，产生矛盾。

(iv) 计数器为零后，每走过一条边必把我们带到一个新的顶点。由于图是有限的，我们一定会回到 v_0。这就完成了证明。

下面的附注很容易验证，其证明从略。

(1) 如果探索迷宫的目标是寻找财宝或怪物米那多，只要目标一找到，就可放弃 Tarry 算法，马上开始按走过的进

入边折回。这通常会较快地返回到入口。

(2) 如果使用一个计数器，当计数器记录增加一个单位时，在原数字上打叉而不擦掉它。最终打叉数的数目与迷宫顶点数目（不计 v_0，但计所有其他交叉点且包括死胡同的终点）相同。

(3) 如果随着游历的进行，交叉点被计数，且走过它们的序列 S 被记录下来，只要计数器一变为零，旅行者马上会知道他在到达入口之前还必须走过的边数 N。数 N 可按如下办法从 S 中得知：设 v_k 为计数器变为零的顶点；设 v_j 是 S 中 v_k 第一次出现时紧挨在 v_k 之前的顶点；v_i 是 S 中 v_j 第一次出现时紧挨在 v_j 之前的顶点，如此等等。那么序列 v_k，$v_j, v_i, \cdots, v_0$ 代表返回的游历路线，这个序列中的顶点个数为 $N+1$。

在第一次游历后，序列 S 也可使人们画出迷宫的一个完整的构图。

例 3 考虑图 4 代表的迷宫及在这个图中用例 1 给出的序列 S 所示的游历。计数器在此序列中最后出现 6 时变为零。紧挨在第一次出现的 6 之前的数为 5，序列 6，5，4，3，2，1，0 代表了这个完整的返回路线，于是 $N=6$。

例 4 同一迷宫的符合算法的另一种走法，在图 6 中由序列 S：

$$0,1,2,5,6,6,4,3,1,3,2,7,5,4,7,$$
$$3,4,6,5,2,1,0$$

所示。

由 7 到达 3 时计数器变为零。因而返回路线由序列3，4，6，5，2，1，0给出，且 $N=6$。

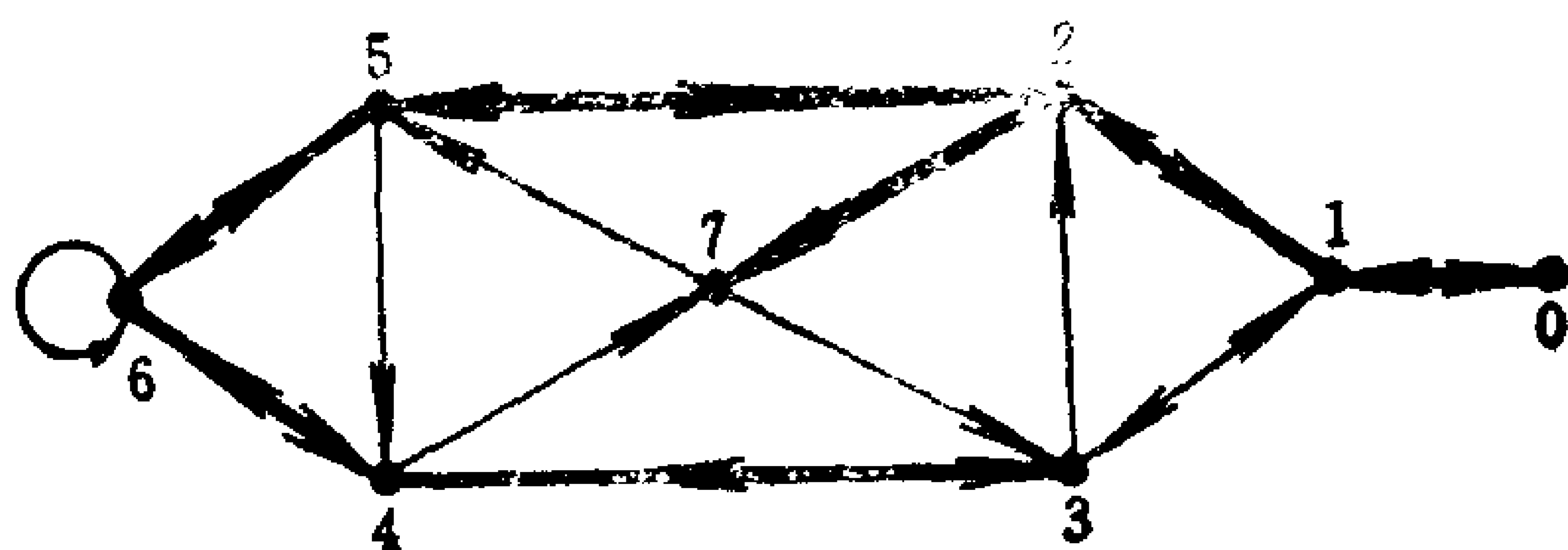


图 6

S的记录有时甚至允许我们缩短从计数器变为零的顶点通向起点的路径．在例 4 中，假定我们沿一条不是它的进入边的边离开顶点 3，我们会到达顶点7,2或 1．从这些顶点的返回路线分别仅含3,2或 1 条边．这样，从 3 返回到起点的路径分别仅包含4,3或 2 条边．

如果除了计数交叉点和记录走过它们的序列之外，还沿迷宫的走廊标记边——在走廊的两端各有一记号——这可得到额外的好处．例如，可在计数器为零之后寻求返回起点的最短路径，仅仅借助于序列 S 可找出一条这样的路径．走廊上的记号可帮助实现这一计划．在例 4 中，可使旅行者从计数器变成零的顶点 3，通过点 1 直接到点 0．

边的标记也可使我们在以后按最经济的路线游历迷宫．

参 考 文 献

[1] W.W.R.Ball, Mathematical Recreations and Essays, rev.by H.S.M.Coxeter, Macmillan, New York, 1947.

[2] D.König, Theorie der endlichen und unendlichen Graphen, Chelsea, New York, 1950.

[3] O.Ore, Theory of graphs, *Amer.Math.Soc.Coll.Publ.*, 38 (1962).
[4] B.A.Trakhtenbrot, Algorithms and Automatic Computing Machines, (transl.), Heath, Boston, 1966.

（刁在筠编译，朱本仁校）

Euclid游戏的性质①

E.L.Spitznagel,Jr.

1. 引言

在最近发表的一篇文章([3])中，Cole 及 David 描述了一种基于 Euclid 算法进行的很有意思的小游戏，并恰如其份地冠之以 Euclid 的名字。他们指出，Euclid 不仅提供了游戏的一个简单例子，而且清晰地阐述了谋胜策略。在[3]中，作者主要是解释谋胜策略，没有谈及 Euclid 策略的其他任何一点令人感兴趣的特性。本文试图在这方面作一些说明。

为完整起见，我们先简要地叙述一下游戏及其谋胜策略，然后再讨论其他的一些性质。

2. 规则和策略

我们按照[3]解释 Euclid 的规则。设(p,q)是一对正整数，$p>q$，并设A和B是两个游戏者。游戏是这样进行的：每个人每次必须走一着，就是在给他的两个数中，用大数减去小数的某个正整数倍得到一个新的非负整数，并用它代替原先的大数。谁能得到 0 作为新的小数，谁就获胜。例如，

① Properties of a game based on Euclid's algorithm, *Math.Magazine*, Mar.-Apr.(1973), 87—92.

最先给定的一对数是（51,30）。则游戏可能如下进行：

		或者		
	(51,30)			(51,30)
A：	(30,21)		A：	(30,21)
B：	(21,9)		B：	(21,9)
A：	(9,3)		A：	(12,9)
B：	(3,0)		B：	(9,3)
B	获胜；		A：	(3,0)
			A	获胜。

这种游戏的谋胜策略可以概括如下：

定义 令

$$c=(\sqrt{5}-1)/2\doteq 0.618$$

（这就是黄金分割）。并设$p>q$，(p,q)是在 Euclid 游戏过程中出现的一对数。如果$q/p>c$，则称(p,q)是安全位置，否则称之为危险位置。

命题1 处在危险位置的游戏者下一步总能走到安全位置。

命题2 处在安全位置的游戏者只能有一种走法①，这种走法的结果永远是到达危险位置。

这两个命题的证明作为习题。

综合上面两个命题可以看到：一旦游戏者A处在某个危险位置上并轮到他走，则他可以确信，以后的每一步，他总能走到安全位置上；相反地，另一个游戏者B，每次都只能被迫走到危险位置上去。因为在游戏过程中，大数p是严格递减的，又因为在游戏开始时$p\neq q$，所以在游戏的某一步

---

① 这时$2q>p>q$，因此只能走到$(q,\ p-q)$。

上必到达这样的危险位置：

$$(kq,q)①, \qquad (1)$$

其中的$k>1$是个整数。因此，游戏者A想获胜的谋略是：每次都走在安全位置上，直到B走到危险位置（1），接下来的一步，A就能得到0从而获胜。

3. 不用黄金分割的谋胜策略

如果一个人根本不知道什么是黄金分割，那么能不能教会他谋胜策略呢？回答是肯定的。对游戏作一番仔细的分析，可得到下面的结论。

命题3 设$p>2q$且p/q不是整数。假定轮到A走，则他至少能走到两个不同的位置。但在所有的走法中，只有一种能使他到达安全位置。若设p_0是p除以q得到的余数，令$p_1=p_0+q$，则安全位置是数对

$$(p_1,q) \text{ 和 } (q,p_0)$$

中，第2项与第1项之比较大的那个数对。

证明 因为$q/p<1/2<c$，所以A准备离开的位置是危险位置。由命题1，下一步，A总能走到安全位置上去。A能走到的位置是

$$(p_1,q), \quad (q,p_0)$$

及所有形式为

$$(p_0+kq,q), \quad p_0+kq<p, \quad k=2,3,\cdots,$$

的数对。当$k\geqslant 2$时，$q/(p_0+kq)<q/kq=1/k<c$，因此，可能的安全位置只剩下(p_1,q)和(q,p_0)。已知其中必有一个

① 若当$q>1$时一直不出现这样的位置，则最终必有$q=1$从而必定如此。

是安全的．下面证明不可能二者均安全．不妨设 (q,p_0) 是安全的，则 $p_0/q>c$，但是

$$\frac{p_1}{q}=\frac{p_0+q}{q}=\frac{p_0}{q}+1>c+1=\frac{1}{c},$$

即 $q/p_1<c$，从而 (p_1,q) 是危险位置．因此，在 (p_1,q) 和 (q,p_0) 中有且只有一个是安全位置，也就是说，在两对数中，有且只有一对数，它的后项与前项的比值大于 c．于是，只要判断出哪对数中，后项与前项的比值大，它就必定是安全位置．

命题 3 得证．

在 Euclid 游戏中，只有处在以下两种类型的位置 (p,q) 时，游戏者才有可能有两种以上的走法供选择：

(1) $p/q\geqslant 2$是一个整数，

(2) $p>2q$ 但 p/q 不是整数．

（想一想，为什么？）而后者正是命题 3 中的假定．

因此，如果不用到黄金分割 c，则谋胜的策略可叙述如下：如果下一步的走法有两种以上的可能性，则

(i) 如果 p/q 是整数，就可得到 0 而获胜；

(ii) 如果 p/q 不是整数，则先计算

$$p_0=p\text{ 除以 }q\text{ 所得余数}$$

及

$$p_1=p_0+q,$$

然后走到 (p_1,q) 与 (q,p_0) 中后项与前项之比较大的那个位置上去．

其实，也用不着先计算比值然后再挑选．如果 $p_0p_1<q^2$，则 $q/p_1>p_0/q$，从而 (p_1,q) 是安全位置，否则 (q,p_0) 是

安全位置。例如，设数对是（70,11）。则可以算得 $p_0=4$ 及 $p_1=15$。因为 $4\cdot 15<11^2$，所以下一步应走到（15,11）而不是（11,4）。

4. 与 Fibonacci 级数的联系

前面，我们已经掌握了 Euclid 游戏的谋胜策略，本节讨论游戏中可能发生的这样的情形：两个游戏者都不知道谋胜策略，而且他们走的每一步都是被迫的"逼着"，即：所处的位置 (p,q) 满足不等式 $q<p<2q$，从而没有选择余地，只能走到位置 $(q,p-q)$。

当然，当某个游戏者所处的位置 (p,q) 满足不等式 $p\geqslant 2q$ 时，这样的一串走法就终结了。在这样的一对数在整个游戏过程中只出现一次的情形，则它必出现在游戏的末尾且形式是（1），从而下一个游戏者得到 0 而获胜。将游戏过程中出现的所有有序对中的每一个数（去掉重复）依相反方向排列，则得到数列

$$q,kq,(k+1)q,(2k+1)q,(3k+2)q,(5k+3)q,\cdots.$$

这是 Fibonacci 数列的 q 倍。Fibonacci 数列在 Euclid 游戏中出现，本不应该感到惊奇，因为它与 Euclid 算法（[4]）及黄金分割（[1],[7]）的联系是熟知的。

现在假设这一连串的"逼着"停止在位置 (p_0,q_0)，$p_0\geqslant 2q_0$ 但不一定取形式（1）。按照游戏进程的反方向返回，我们又会看到 Fibonacci 数列。设在 (p_0,q_0) 之前第 i 步上的数对是 (p_i,q_i)，则有递归表达式

$$(p_i,q_i)=(p_{i-1}+q_{i-1},p_{i-1}). \qquad (2)$$

对这些数对我们有以下的结论，它十分类似于大家熟知的有

关连分数的渐近分数的结果（[2],[5],[7]）。

定理 给定 (p_0,q_0)，设 (p_i,q_i) 由递归关系（2）定义，则有

(i) q_{2k}/p_{2k} 是 k 的单调递增函数，$k=0,1,2,\cdots$；

(ii) q_{2k+1}/p_{2k+1} 是 k 的单调递减函数，$k=0,1,2,\cdots$。

证明 由(2)得

$$q_{i+1}p_i-p_{i+1}q_i=p_i(p_{i-1}+q_{i-1})-(p_i+q_i)p_{i-1}$$
$$=-(q_ip_{i-1}-p_iq_{i-1}).$$

重复这一结果 i 次得

$$q_{i+1}p_i-p_{i+1}q_i=(-1)^i(q_1p_0-p_1q_0). \tag{3}$$

记 $d=q_1p_0-p_1q_0$。由（3）式得

$$\frac{q_{i+1}}{p_{i+1}}-\frac{q_i}{p_i}=(-1)^i d\frac{1}{p_{i+1}p_i},$$

$$\frac{q_{i+2}}{p_{i+2}}-\frac{q_{i+1}}{p_{i+1}}=(-1)^{i+1} d\frac{1}{p_{i+2}p_{i+1}}.$$

从而得

$$\frac{q_{i+2}}{p_{i+2}}-\frac{q_i}{p_i}=(-1)^i d\left(\frac{1}{p_{i+1}p_i}-\frac{1}{p_{i+2}p_{i+1}}\right).$$

因为$p_{i+2}>p_i$，所以

$$\frac{1}{p_{i+1}p_i}-\frac{1}{p_{i+2}p_{i+1}}>0.$$

又因为$p_0\geqslant 2q_0$，所以

$$d=q_1p_0-p_1q_0=p_0^2-(p_0+q_0)q_0$$
$$\geqslant 2p_0q_0-p_0q_0-q_0^2=q_0(p_0-q_0)$$
$$>0.$$

因此，若 i 是偶数，则有

$$\frac{q_{i+2}}{p_{i+2}} > \frac{q_i}{p_i};$$

若 i 是奇数，则不等号反向。

定理得证。

于是，回到游戏上来，对于任意一串"逼着"，我们可以有如下的看法：如果一旦在游戏者 A 面前的两个数 (p,q) 满足不等式 $q/p<c$，则他可以发现，在以后的游戏进程中，摆在他面前的两个数之比 (q/p) 会逐渐减小，直至最后给他的一对数之比 $\leqslant 1/2$（见定理），因而使他可以作出一种如何走法的判断。如果游戏者 A 总是根据这个策略来作出他的判断，则游戏者 B 就会发现自己完完全全地陷入了对方的圈套。这就是，在由"逼着"构成的游戏进程中，B 发现：A 交给他的两个数之比 q/p 稳步地上升到 1，但同时他被迫交给对方的两个数之比 q/p 却稳步减小直至比 1/2 小①。然后，A 可以选择这样的一步走法——它可能使得交给 B 的两个数之比 (q/p) 小于他上次交给 B 的两个数之比，但这一定又立即开始了新的一串"逼着"。B 又一次发现，当这下一轮"逼着"开始后，交给他的两个数之比 q/p 又朝着错误的方向前进，即逐渐增大到 1。因此他永不可能有获胜的机会。

与 Nim 游戏② 相比，Euclid 游戏的谋胜策略太不隐蔽了。在 Nim 游戏中，在游戏进程的每一步上（当然，最后一步除外），按照谋胜策略走步的游戏者的对手，都有不止一种走法可供选择，因此没有明显的征兆预示他必然会失败，而在 Euclid 游戏中，按照谋胜策略走步的游戏者的对手，很

① 至此，这一串"逼着"就结束了.

② 参见本丛书的第一册《等周问题和夫妇入座问题》.

容易看出，他走的每一步都是不由自主的，因而也许就会猜出什么是谋胜的策略。

5. 出发点是危险位置的概率

至此，我们已解释了 Euclid 游戏中的谋胜策略。一个自然的问题是：任意选取一个位置作为出发点，它是危险位置的概率有多大？

首先，我们给出这一问题的一种合理解释：设 S 是由小于或等于某个正整数 n 的正整数组成的集合，表示位置的两个数都选自集合 S，其中第一个数 x 的选取是随意的，S 中的每一个元素被选到的可能性是相同的，但第二个数的选取，对于集合 $S-\{x\}$ 中的每一个元素并不都是等可能的，记被选出的两个数分别为 p 和 q 满足 $p>q$。

根据上述解释，由经典的概率理论（见[6]），可以证明 q 是从集合 $\{1,2,\cdots,p-1\}$ 中随机选取的，其中的每一个数被选取的可能性都相等。给定 p，我们有

$$P(q/p<c)=P(q<cp)=[cp]/(p-1),$$

其中[·]表示取整数部分。如果 p 很大，则上述分数非常接近 c。因此对大数 n，随机选到的出发位置是危险位置的概率近似等于 c。这就是说，在 Euclid 游戏中，先走的人，即游戏者 A，能按谋胜策略前进的可能性比较大。

参 考 文 献

[1] W.W.R.Ball, Mathematical Recreation and Essays, Chapter 2, Macmillan, New York, 1960.

[2] G.Chrystal, Textbook of Algebra, Chapter 32, Dover, New York, 1961.

[3] A.J.Cole and A. J. T. Davie, A game based on the Euclidean algorithm and a winning strategy for it, *Math. Gaz.*, 53 (1969), 354--357.
[4] J.D.Dixon, A simple estimates for the number of steps in the Euclidean algorithm, *Amer. Math. Monthly*, 78 (1971), 374—376.
[5] A.Ya.Khinchin, Continued Fractions, Chapter I, University of Chicago Press, Chicago, 1964.
[6] E Parzen, Modern Probability Theory and Its Applications, Wiley, New York, 1960.
[7] N. N. Vorob'ev, Fibonacci Numbers, Blaisdell, New York, 1961.

(朱学贤编译，潘承彪校)

炮眼问题①

Bill Sands

如果把一些2×1的骨牌放在一个$m\times n$板上(每一个骨牌恰好覆盖板上的两个方块)，直到没有更多的骨牌能被容纳为止。这样，就可能有若干个1×1的方块是空的②，我们称之为“洞”。

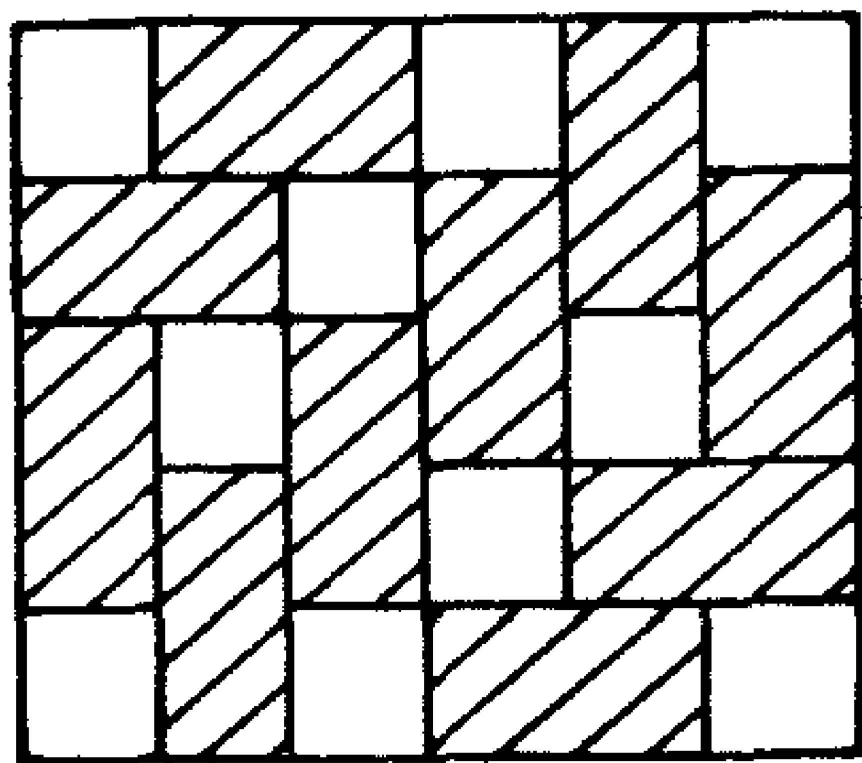


图 1

在上面的例子(见图1)中，我们在5×6的板上放了10个骨牌且留下了10个洞。一个很自然的问题是对于可能留下的洞的数目找一个下界。因为这个问题在城堡的建筑中有明显的实际应用，作者称它为“炮眼问题”。我们将证明下面的命题。

① The gunport problem, *Math. Magazine*, Sept.-Oct. (1971), 193—196。

② 按放骨牌的要求，不可能有1×2或2×1的方块是空的。——译注

命题　对于$m,n>1$，洞的数目≤骨牌的数目。

我们将首先处理两种特殊情况。

(1) 假定$m=1$，则易见对$n\equiv1(\mathrm{mod}\ 3)$命题可能不成立，而对$n\equiv0$或$2(\mathrm{mod}\ 3)$命题成立。$n\equiv1(\mathrm{mod}\ 3)$时命题不成立是由于可以使洞和骨牌交错出现而使板的两端是洞。这样洞比骨牌的数目多1。

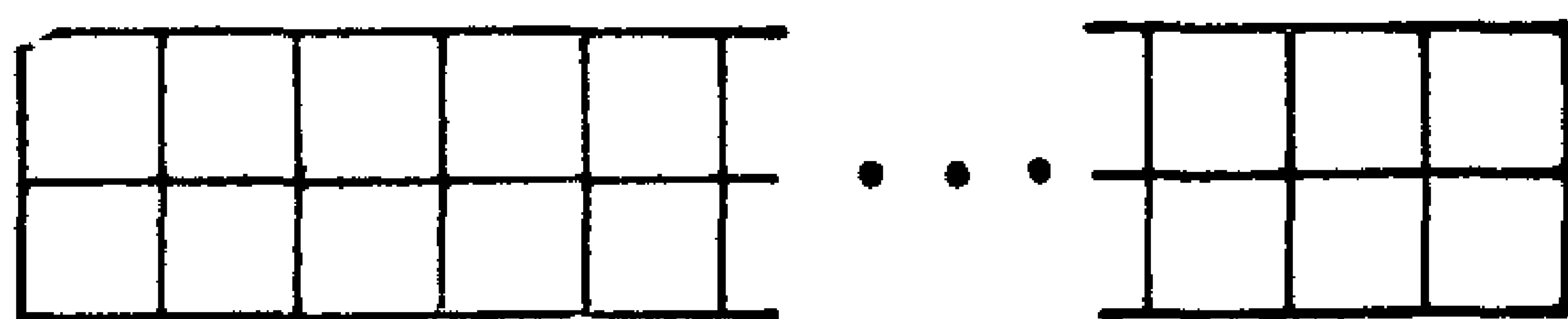


图　2

(2) 假设$m=2$(见图2)，可以用下面的方法证明命题对所有的n都成立：考虑所有竖直放置的骨牌，它们将板分成一些可分别考虑的$2\times t_i$的较小的板。这样我们只需对骨牌都是水平放置的板来证明命题即可。考虑如图3中所示形

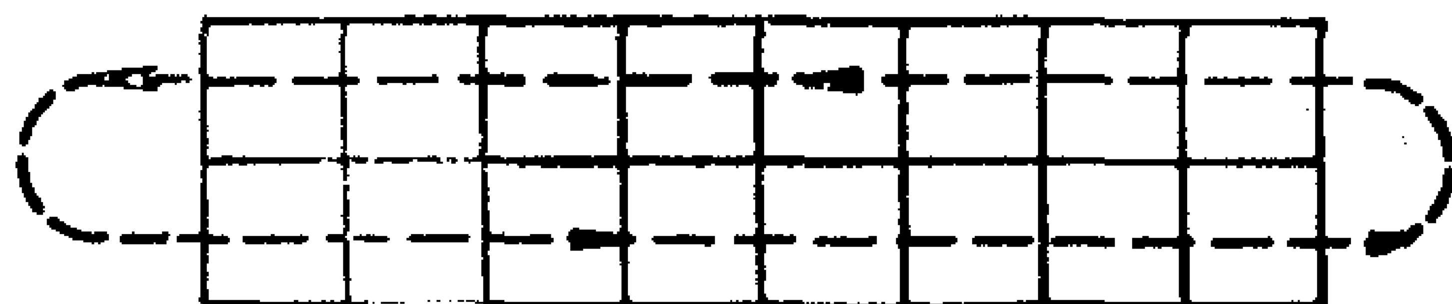


图　3

成一个闭环的洞和骨牌。我们看到洞最多的情况是洞与骨牌交错。因此洞的数目≤骨牌的数目。

这样只需讨论$n\geq3$，$m\geq3$的情况。

我们将下列类型的洞和骨牌标号(参见图4)。

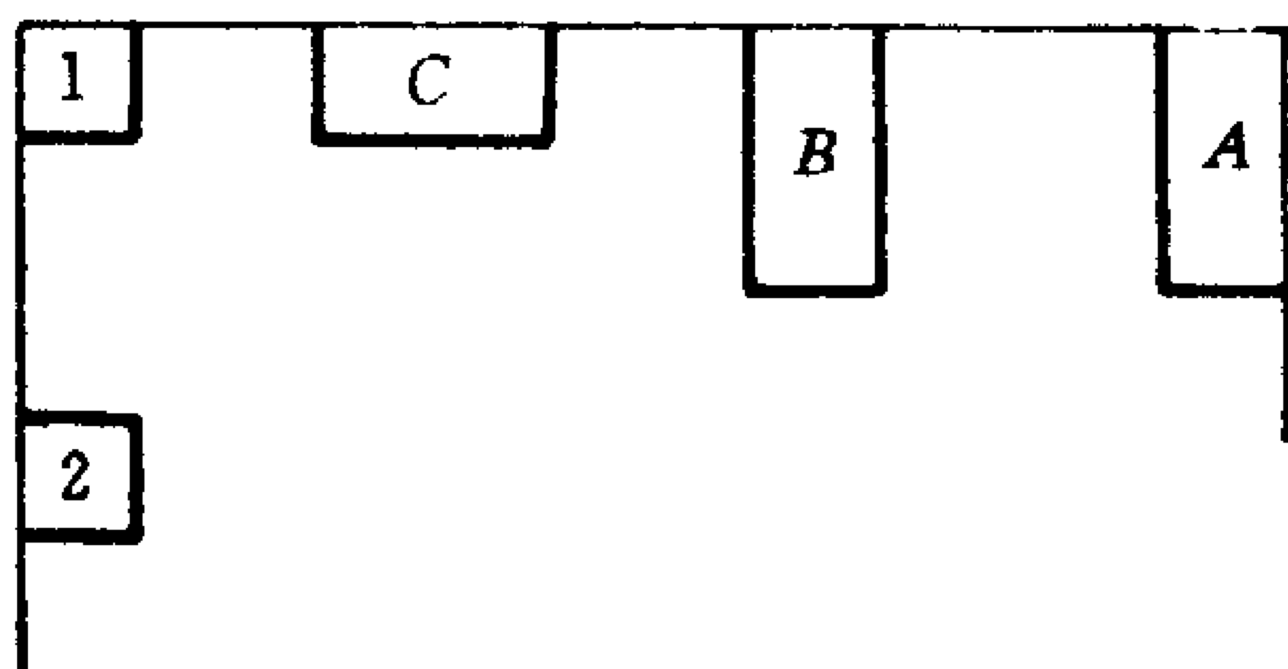


图 4

洞：

1 型——在角上。

2 型——在边上但不是 1 型。

骨牌：

A 型——在角上。

B 型——短边在边上但不是 A 型。

C 型——长边在边上但不是 A 型。

我们用h_i表示 i 型洞的数目，$i=1,2$，而用 d_j表示 j 型的骨牌的数目，$j=A,B,C$。设洞的总数是 h，骨牌的总数是d。于是内洞的数目 $=h-h_1-h_2$，而内骨牌的数 目 $=d-d_A-d_B-d_C$。我们要证明$h\leqslant d$。为此我们将数出不同的对(H,D)的数目，其中H是洞，而D是沿着它的一边与洞接触的一个骨牌。

由骨牌的放法知下面的结论是显然的。

(a) 每个 1 型洞有两个骨牌和它接触。

(b) 每个 2 型洞有 3 个骨牌和它接触。

(c) 每个内洞有 4 个骨牌和它接触。

于是(H,D)对的数目恰好是

$$4(h-h_1-h_2)+3h_2+2h_1=4h-2h_1-h_2$$

而另一方面，

(a) 每个 A 型骨牌至多有 2 个洞与它接触。

(b) 每个 B 型骨牌至多有 3 个洞与它接触。

(c) 每个 C 型骨牌至多有 3 个洞与它接触。

(d) 每个内骨牌至多有 4 个洞与它接触。

于是(H,D)对的最大可能的数目是

$$4(d-d_A-d_B-d_C)+3d_C+3d_B+2d_A$$
$$=4d-2d_A-d_B-d_C.$$

这样我们有

$$4h-2h_1-h_2\leqslant 4d-2d_A-d_B-d_C,$$

或者

$$4h+2d_A+d_B+d_C\leqslant 4d+2h_1+h_2. \qquad (1)$$

因为沿着板边的洞至多可能是与骨牌交错出现，所以，我们知道在板边的洞的数目不超过在边上的骨牌的数目，即

$$h_1+h_2\leqslant d_A+d_B+d_C.$$

于是从(1)我们有

$$4h+d_A\leqslant 4d+h_1. \qquad (2)$$

由于显然有 $d_A+h_1=4$，于是有

$$2h+d_A\leqslant 2d+2.$$

这样如果$d_A=2,3$或 4，我们得到$h\leqslant d$。

又若 $d_A=1$，则$h\leqslant d$，这是因为 h 和 d 是整数。最后讨论$d_A=0$的情形，即每个角上有一个洞。这时有

$$h\leqslant d+1.$$

我们来证明这里等号不能成立，即必有$h\leqslant d$。假定$h=d+1$，则上面所有的不等式变为等式。特别地(2)和(1)变成等式。因此每个骨牌都必须有最大数目的洞与它接触。我们来考虑

板的一个角。由于假定每个角上有一个洞，所以只可能有两种方法绕着这个角上的洞放骨牌，如图 5 所示。

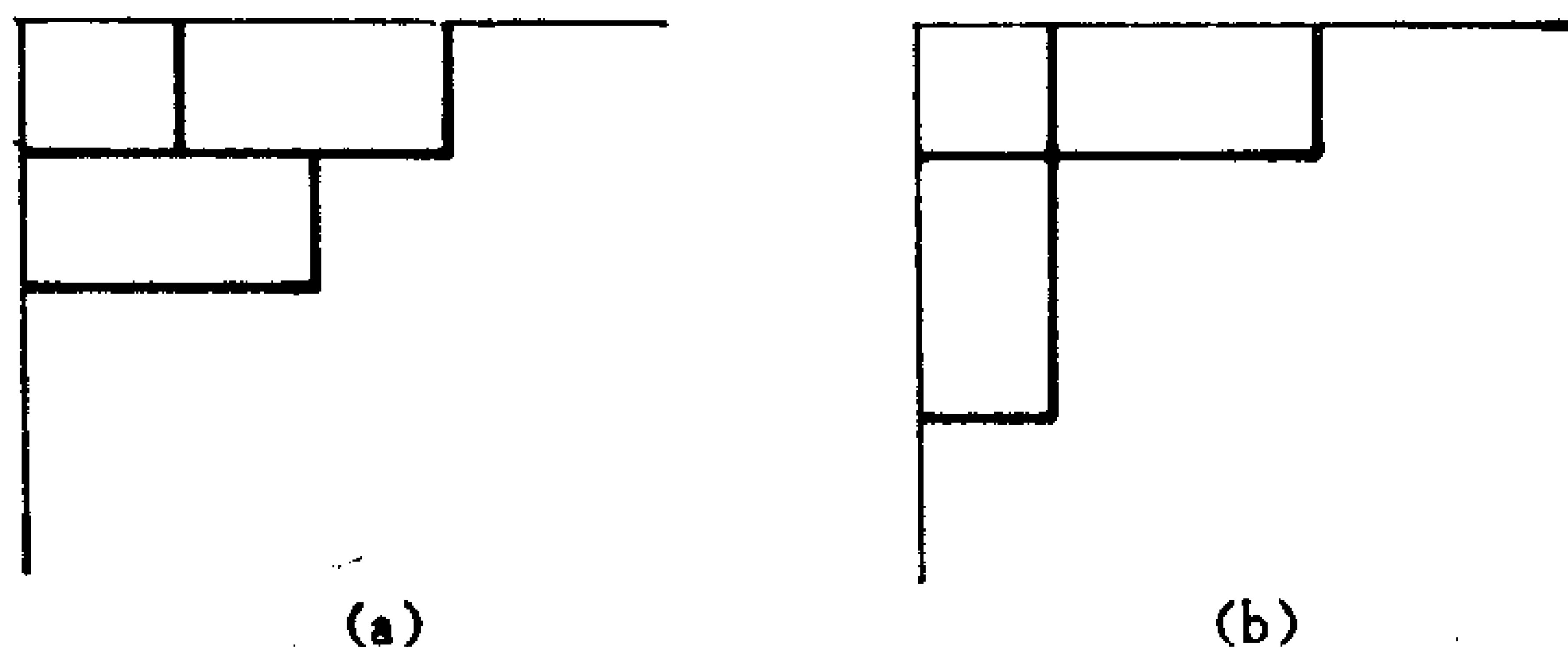


图 5

读者可以检验：由每个骨牌必有最大数目的洞与它接触的假定，从图 5 的(a)和(b)分别唯一地产生下面的重复的图案，即图6(a)和图6(b)。

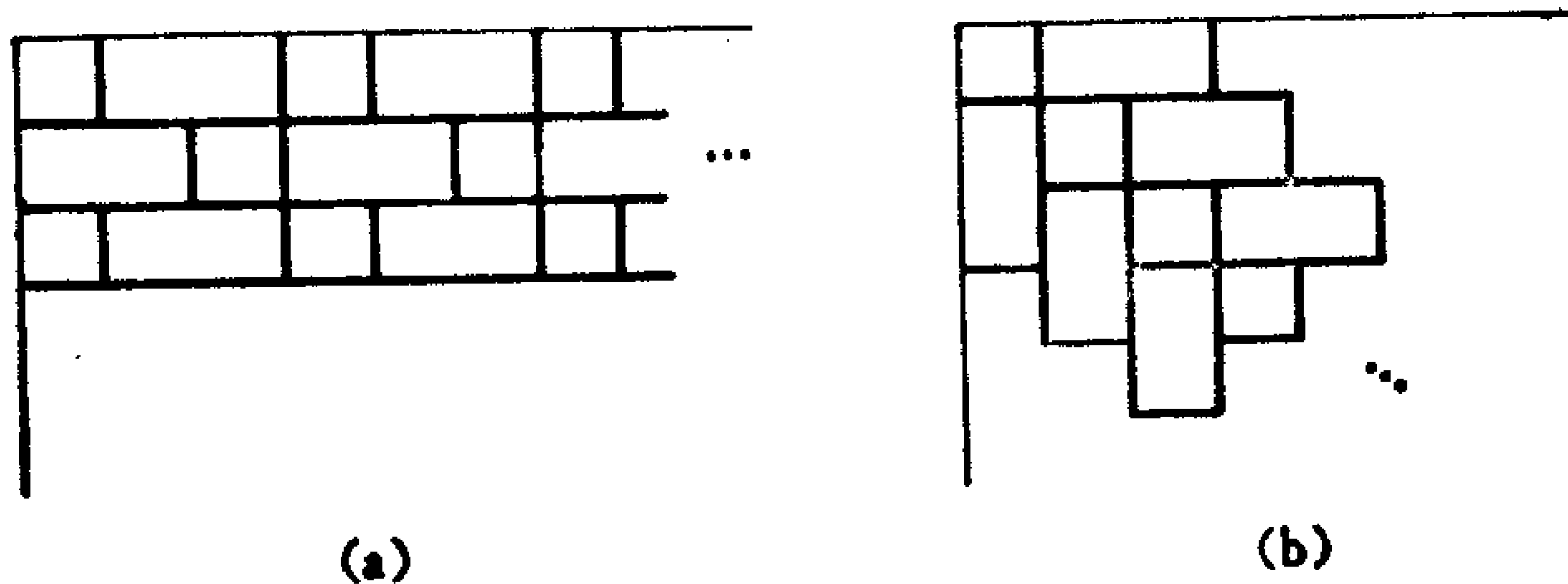


图 6

在图6(a)中，在板的右上角不可能有洞。

在图6(b)中，没有长方形的板能用这种图案填满。

这就证明了$h=d+1$不可能成立。这样对d_A的所有值 $h\leqslant d$ 且命题成立。

注意：若 n 或 m 是 3 的倍数，则洞的数目可能等于骨牌

的数目，考察上面的图6(a)易见这是显然的。否则当然洞的数目将总是小于骨牌的数目。当$\min(m,n)\leqslant 7$或$m,n\leqslant 9$时，能够找到例子使其达到最大可能的洞的数目①。作者至今没有能找到这样一个具体例子的最小板是8×10板。已经证明它可能含有的洞的最大数目是26个(具有27个骨牌)；可是至今所找到的最好的例子是有24个洞，28个骨牌。

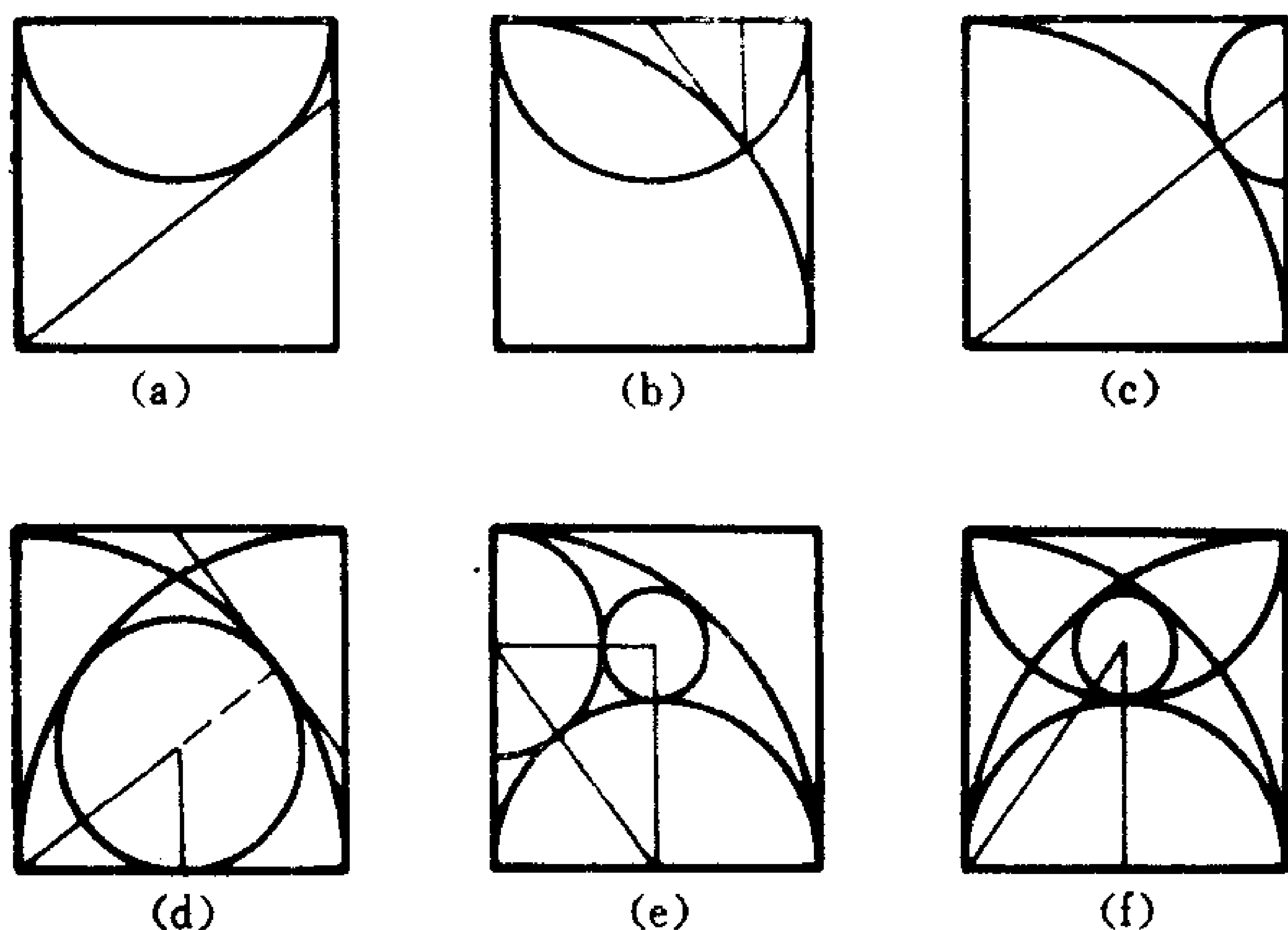
(刘桂真译，潘承彪校)

① 注意："最大可能"的洞的数目不一定就是和骨牌数相等，可以小于骨牌数——译注。

无处不在的3:4:5三角形①

L.Bankoff, C.W.Trigg

在研究与正方形连结在一起的圆弧和圆之间的关系时，我们对经常不期而至的 3:4:5 三角形感到好奇，其中的一些展示在图 1 的正方形中。古代Pythagoras学派为之感到欣喜，



(a) (b) (c)

(d) (e) (f)

图 1

因为这给他们的数的神秘主义提供了明显的佐证。其实一点

① The Ubiquitous 3:4:5 Triangle, *Math. Magazine*, Mar.-Apr. (1974), 61—70.

也不神秘，我们已经看到和找到了下述的产生 3:4:5 三角形的某些构型。但是，仍有许多这种三角形是在孤立的情形中出现的。

1. 线性网格

图1(a)是图 2 开始时的样子——一个 正 方 形 $ABCD$，DC边上有一个半圆(F)，它的切线AG的延长线交 BC于H．(在本文的整个讨论过程中，为简便起见，圆和圆弧都用它们的圆心字母放在括号里表示。)为计算简便，正方形边长取为24，因此 $DF=FC=FG=12$。

连结 AB 的中点E及 CD 的中点F，分别交半圆(F)及

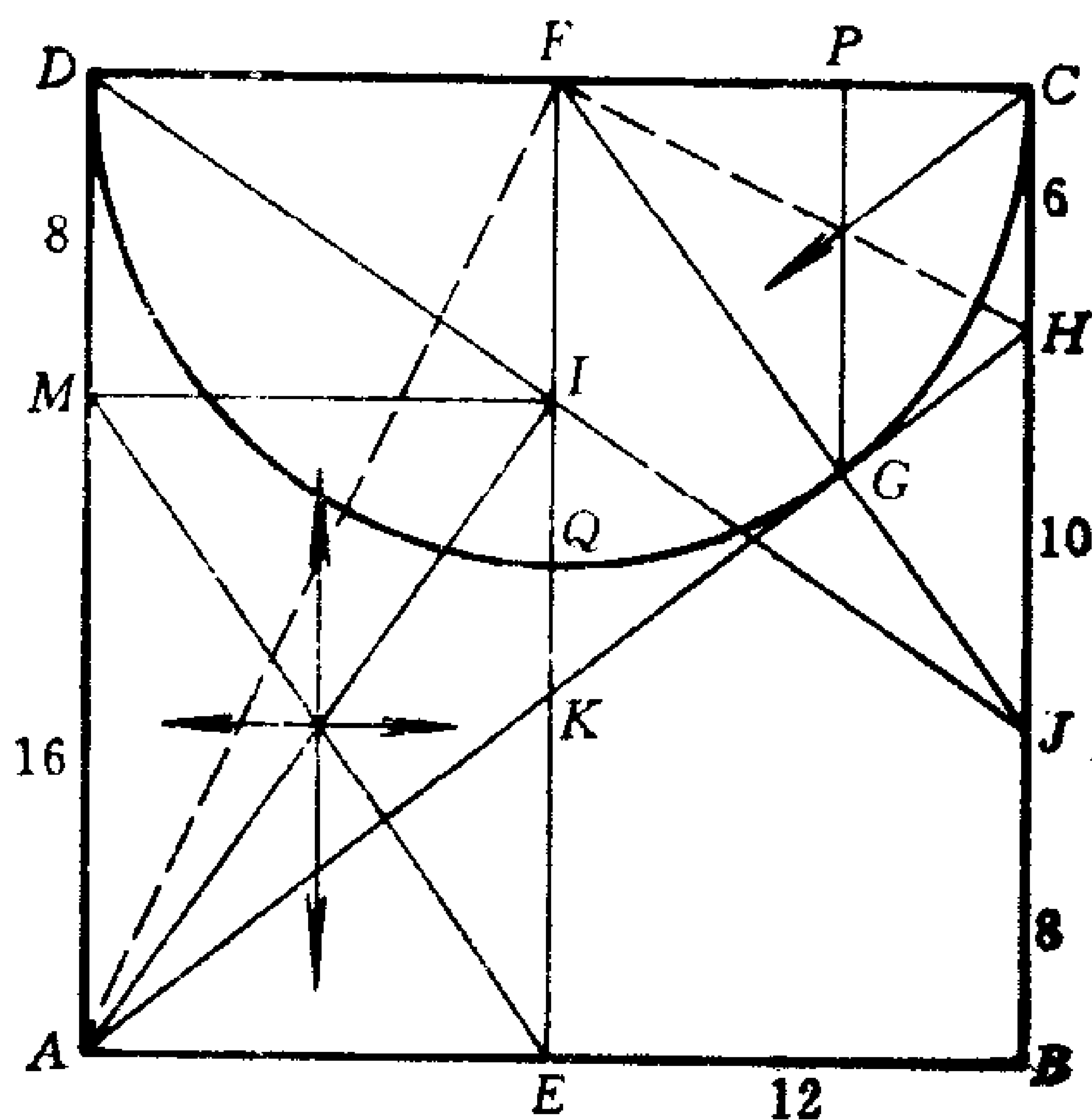


图 2

AH于Q和K．半径FG的延长线交CB于J且与AH相垂直．切

线$HG=HC$，切线$AG=AD=24$．直角三角形FHG与直角三角形FHC全等，直角三角形FAG也全等于直角三角形FAD，因而斜边FA和FH分别平分$\angle GFC$及$\angle GFD$。于是$\triangle AFH$是直角三角形，$HG=(FG)^2/AG=(12)^2/24=6=HC$。

因此有：$HB=18$，$AH=30$，$AK=15$，$KG=9=KE$，$FK=15$及$QK=3$。因为$\triangle FGK\backsim\triangle JGH$，所以有$HJ=10$，$JB=8$，$GJ=8$，$FJ=20$及$CJ=16$。$DJ$交$FE$于$I$，因此$FI=8$，$IQ=4$。

于是，$\triangle FCJ$，$\triangle HGJ$，$\triangle FGK$，$\triangle AKE$和$\triangle ABH$(图1(a))都是3:4:5三角形。

还可注意到，正方形的边长也被分成长度比为4:5:3的3段。

过点G作BC的平行线交FC于P。由于$\triangle FCJ$与$\triangle FPG$相似，因此$\triangle FPG$也是3:4:5三角形。

过点I作AB的平行线交DA于M，得到边长分别是12和16的矩形$IMAE$。从而可得：$\triangle EMI$和$\triangle AIE$都是3:4:5三角形且$ME=20=AI$。

显然，这一网格可以加细从而得到任意多个3:4:5三角形，只要去作那些已经确定的边的平行线或垂直线(图2中箭头所示）就可以了。或者，将图形关于EF反射也可得更多的这种三角形。

2．相关的圆网格

在图3的圆网格中，基本的点如图2一样。其中1/4圆(A),(B),(D)的半径都是24；半圆(E),(F)的半径都是12；半圆(M)的半径是8；半圆(H)的半径是6；圆(K),(I)的

半径是4。在前一节中得到的长度和角在下面要用到。

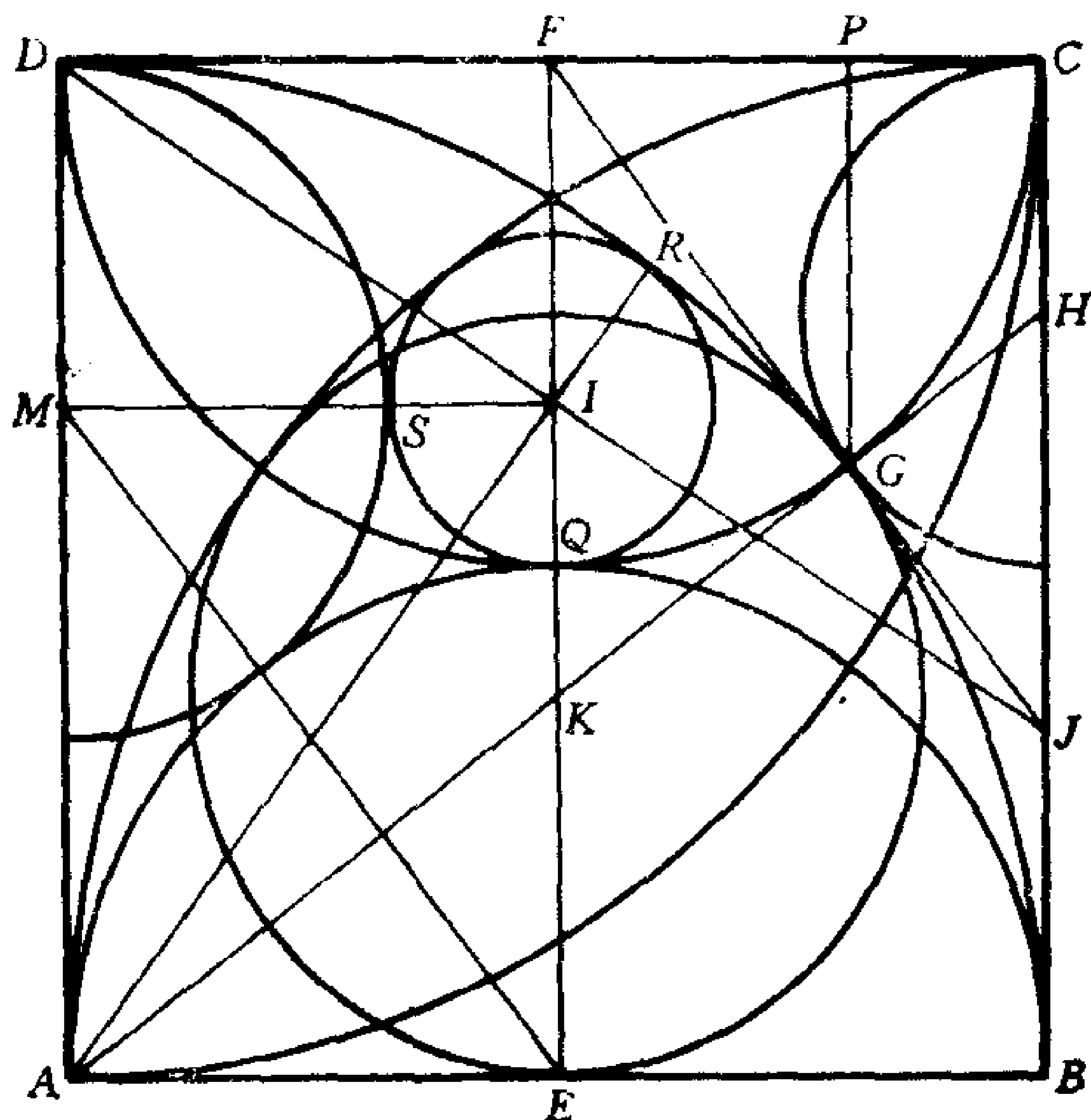


图 3

因为 $DA=AG=AB=24$，所以 G 是圆 (A) 与 (F) 的交点（图1(b)），$\triangle FPG$ 是3:4:5三角形。

因为 $HG=HC=6$，$AH\perp FJ$，所以圆 (A) 和 (H) 相切（图1(c)），$\triangle AHB$ 是3:4:5三角形。

A,K 和 G 三点共线。$KE=KG$，从而圆 (K) 与 AB 相切且与圆 (A) 相切，由对称性，也与圆 (B) 相切（见图1(d)）。$\triangle AKE$ 及 $\triangle FCJ$ 都是3:4:5三角形。

因为 $ME=20=8+12$，所以圆 (M) 和 (E) 相切。$IS=IM-SM=12-8=4=IQ$，$IR=AR-AI=24-20=4$，因而圆 (I) 与 (M)，(A)，(B)，(E) 及 (F) 都相切（图1(e)，(f)）。$\triangle EMI$ 和 $\triangle AIE$ 都是3:4:5三角形。

3. 3张有关的图形

在图 4 中，要求在由(E)，(B)及 CB 围成的曲边三角形

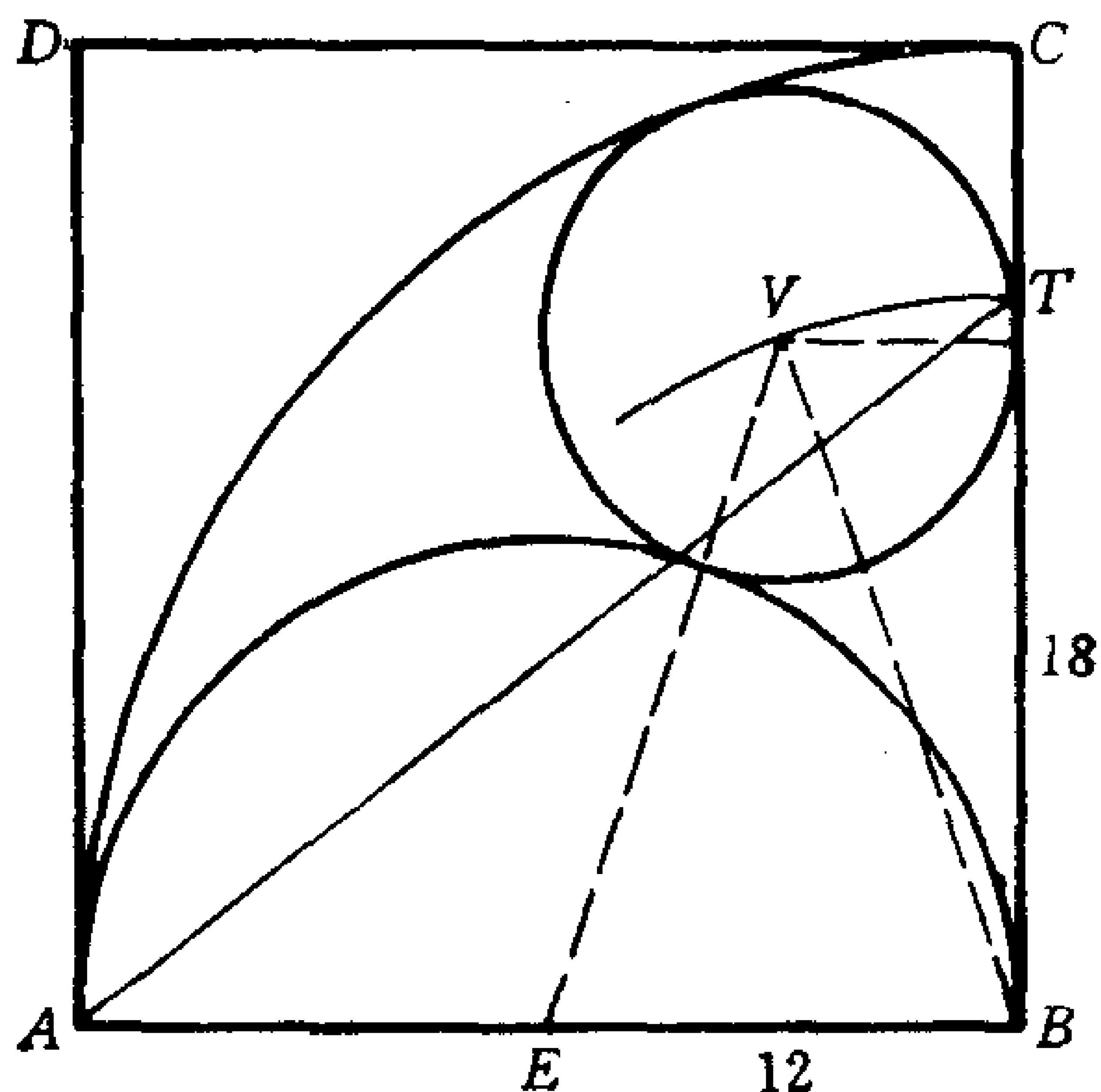


图 4

中作一个内切圆(V)．如果内切圆的半径是 r，则$VE=12+r$ 及 $VB=24-r$。如果二者相等，则 $r=6$，这也是点V到 BC 的距离。要作的内切圆的圆心V位于分别以E和B为圆心，半径都是18 的两个圆的交线上，这两个圆中的后一个交 BC 于T，因此$\triangle ABT$是3:4:5三角形。

将图 3 中的圆(A)，(H)和(I)分离出来，画在图 5 中。延长AI 与(I)的平行于AD 的切线相交，则得到与$\triangle AIE$ 相似的3:4:5三角形 AXY。因为$AY=12+4=16$，所以 $XY=64/3$，$AX=80/3$。又得 $ZH=CH=6$。于是，$RX=8/3=WX$及 $XZ=XH-ZH=\sqrt{8^2+(6-8/3)^2}-6=8/3$。因此 X是与(A)，(H)及DC相切的圆的圆心。

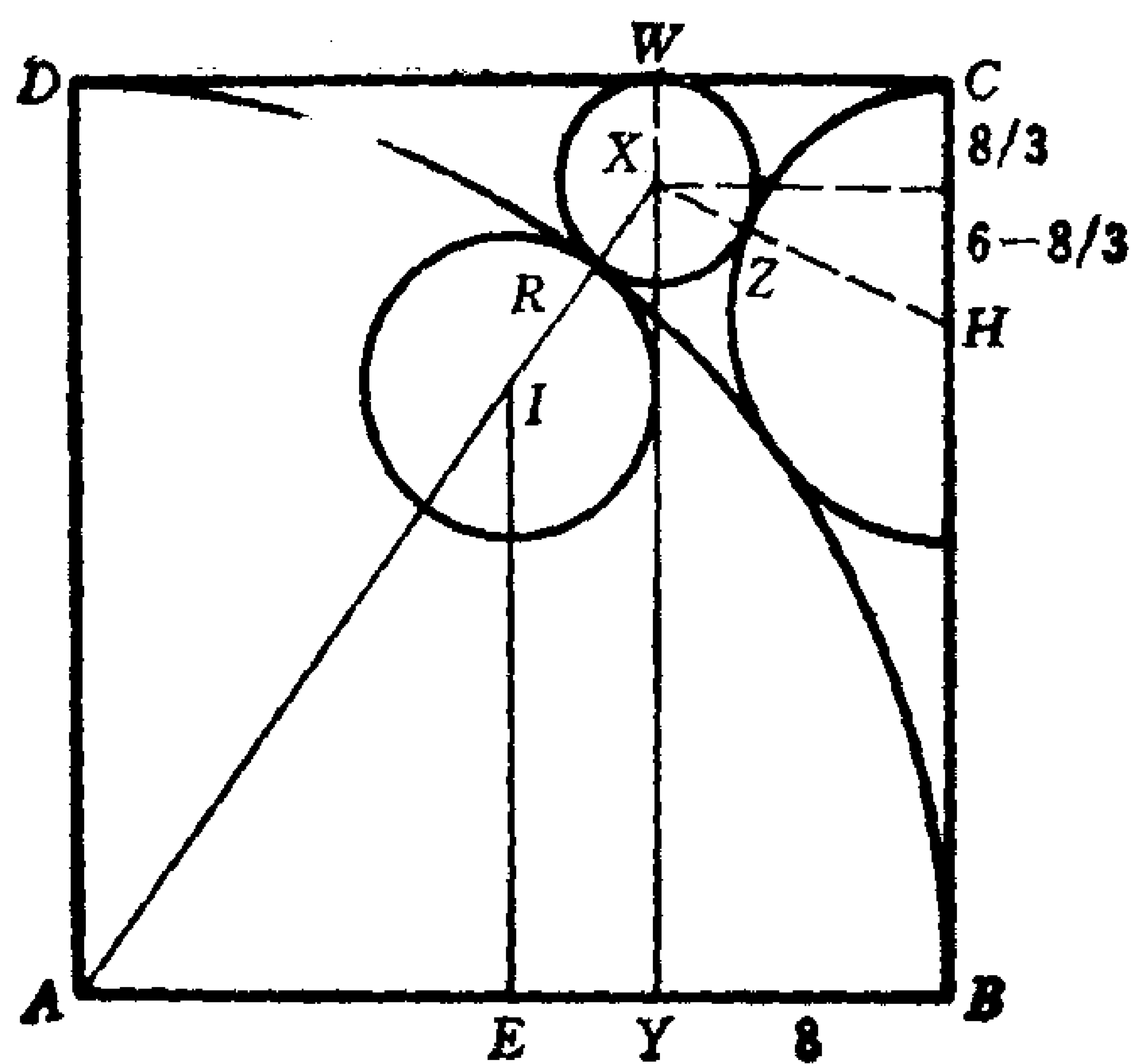


图 5

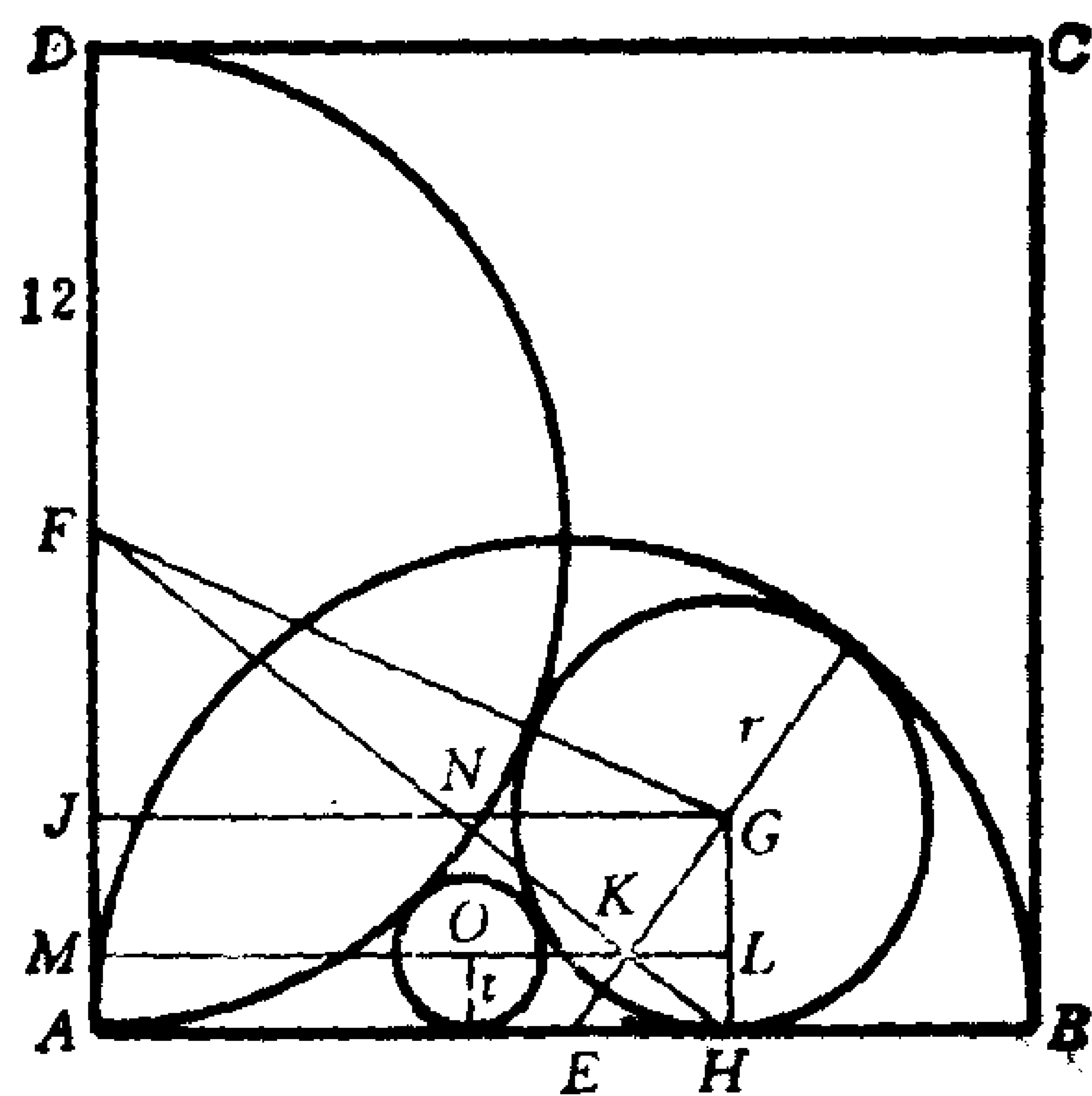


图 6

在图 6 中，半径为 r 的圆（G）内切于以 AB，圆（E）及（F）为边的曲边三角形，（E）和（F）的半径都等于12。可以

看到

$$(FJ)^2+(JG)^2=(FG)^2, \qquad (EH)^2+(GH)^2=(EG)^2.$$

令$EH=x$，则上两式为

$$(12-r)^2+(12+x)^2=(12+r)^2, \qquad x^2+r^2=(12-r)^2。$$

由这 2 个方程可得

$$x^2+8x-48=0。$$

它的正根是$x=4$。因此$r=16/3$。

由此推得$\triangle EGH$，$\triangle FHA$，$\triangle FNJ$及$\triangle HNG$都是3:4:5三角形，且$FH\perp EG$，垂足是K。从各直角的顶点出发，作那些已经确定位置的斜边的垂线，所得图形就是3:4:5三角形的一个极为丰富的网格。在图6中，$\triangle GHK$，$\triangle EHK$，$\triangle KGL$及$\triangle KHL$($KH=16/5$，$LH=48/25$)都是3:4:5三角形。

若一个圆与另两个圆相切并与它们的公切线也相切，则它的半径的平方根的倒数等于那两个圆的半径的平方根的倒数之和([1])。因此，如果设由AH，圆(F)和(G)所围的曲边三角形的内接圆半径为t，则有$1/\sqrt{t}=1/\sqrt{12}+1/\sqrt{16/3}$，解得$t=48/25=LH$。

4. 4个以上的图形

在图7中，(P)和(Q)是两个半径相等的圆，它们的圆心分别落在另一个圆的圆周上。DE是两圆圆心连线的延长线。AB是这两个圆的公弦且是圆(P)，(Q)的根轴；圆(F)是由弧QA，AE及线段EQ所围曲边三角形的内切圆。(G)是由(P)，(Q)及(F)所围的较小的那个曲边三角形的内切圆。

如果一个动圆与两个定圆相切，则动圆的半径长与其圆

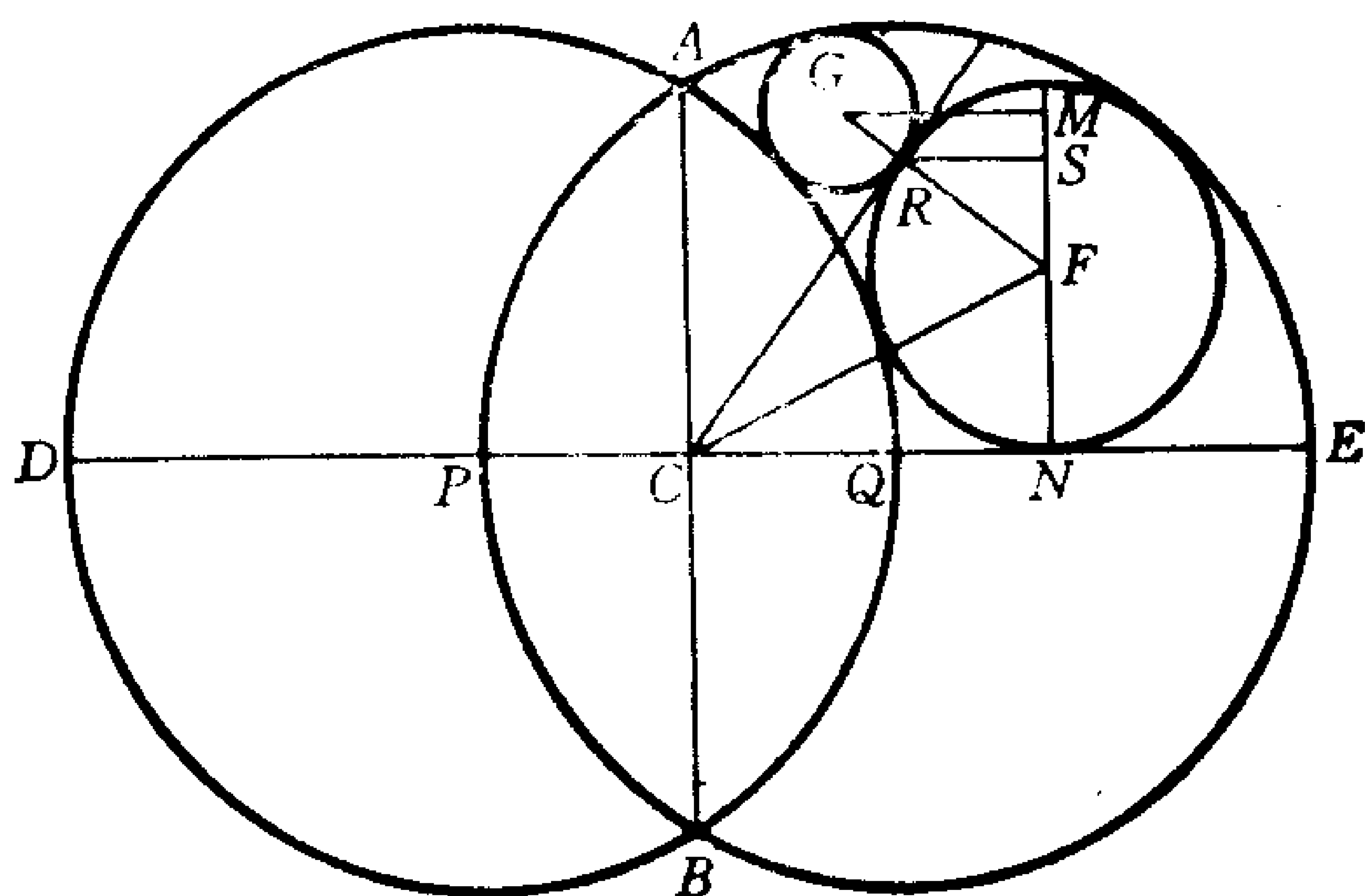


图 7

心到那两个圆根轴的距离的比是一个常数([2])。现在，(G)和(F)是动圆的两个位置。半径最大的动圆是以 QE 为直径，因此常数比是1:2而且$CN=2FN$。

如果两个圆都与另两个圆相切，而且都属于同一种类型，则任意一对的根轴都通过相应的另一对的相似中心([3])。C 是(P)和(Q)的内相似中心，因此落在(G)和(F)的内公切线CR上；$CR\perp GF$；$FN\perp CN$，作GM垂直于 NF的延长线。$CR=CN$，因此 $\angle MFG=\angle NCR=2\angle NCF$。$\tan\angle NCF=1/2$，$\tan\angle MFG=\tan\angle NCR=4/3$，于是$\triangle FMG$是3:4:5三角形，$\triangle FSR$也是。

引理 在直角三角形中 $a^2+b^2=c^2$，如果 $c-b=b-a$，则其边长之比是3:4:5。

证明 因为 $c=2b-a$，所以 $a^2+b^2=4b^2-4ab+a^2$，或者 $b(3b-4a)=0$。因此 $a=3k,b=4k,c=5k$。

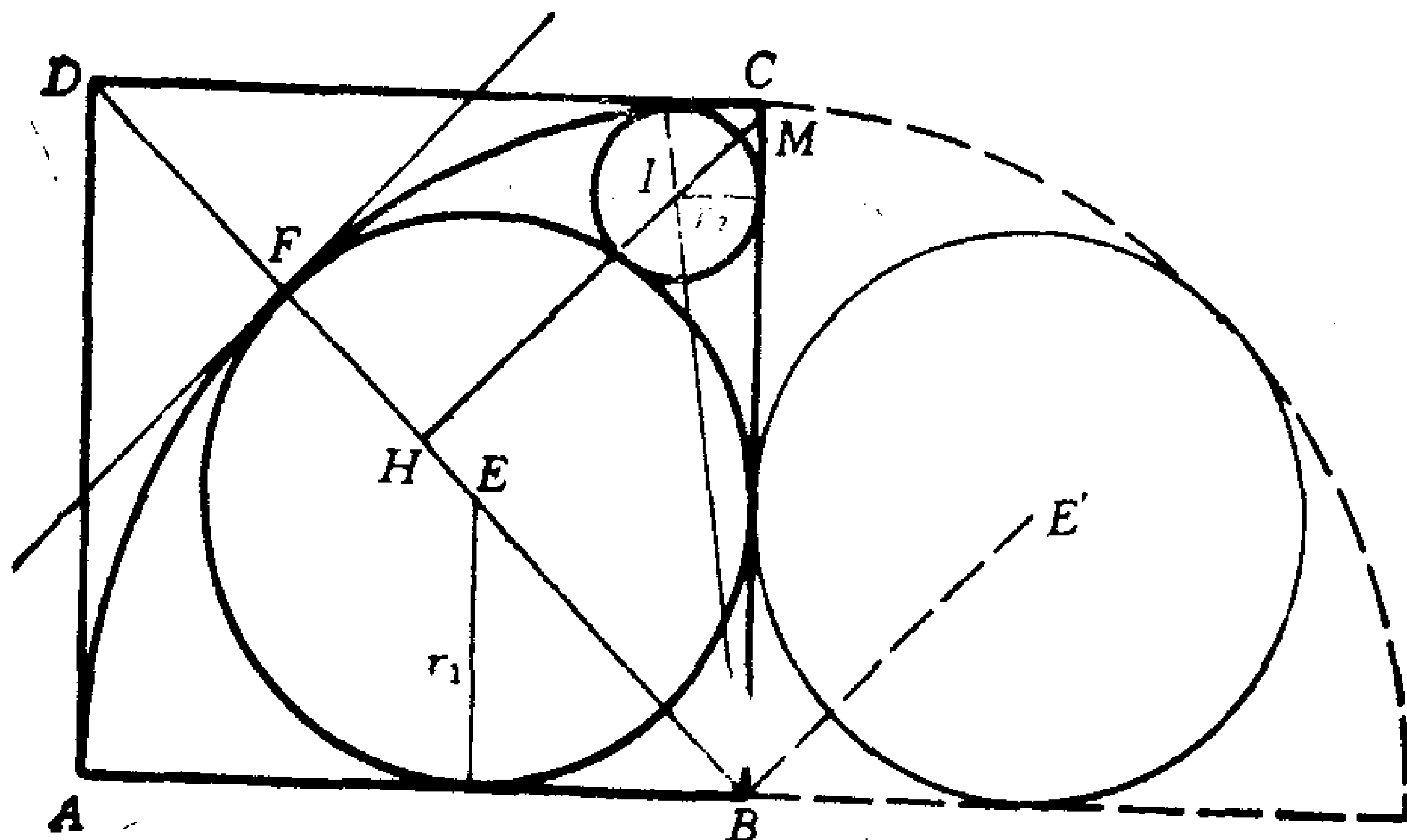


图 8

在图 8 中，半径为R的1/4圆（B）内切于正方形$ABCD$。半径为r_1的圆(E)内切于1/4 圆(B)；半径为r_2的圆(I)和(E),(B)及CB都相切。圆心B,E的连线延长交圆(B)于F。过F作BF的垂线,则它与圆(B),(E)相切且是它们的根轴。作$HI\perp FB$，其延长线交CB于M；再关于BC反射(B)和(E)。

将(E')及(I)看成是动圆的两个位置，并应用前面提到的 Casey 的定理([2])，则有$r_1:FB=r_2:FH$。因为$R=FB=FE+EB=r_1(1+\sqrt{2})$，所以$FH=FB(r_2/r_1)=r_2(1+\sqrt{2})$。进一步有$HB-HI=HM-HI=IM=r_2\sqrt{2}$。还有$IB-HB=(R-r_2)-[R-r_2(1+\sqrt{2})]=r_2\sqrt{2}=HB-HI$。因此由引理得，$\triangle IBH$是3:4:5三角形。

不失一般性，图 9 中正方形$ABCD$的边长取成1，它也是1/4圆(D)的半径长。则1/4圆(B)的半径等于$\sqrt{2}-1$。

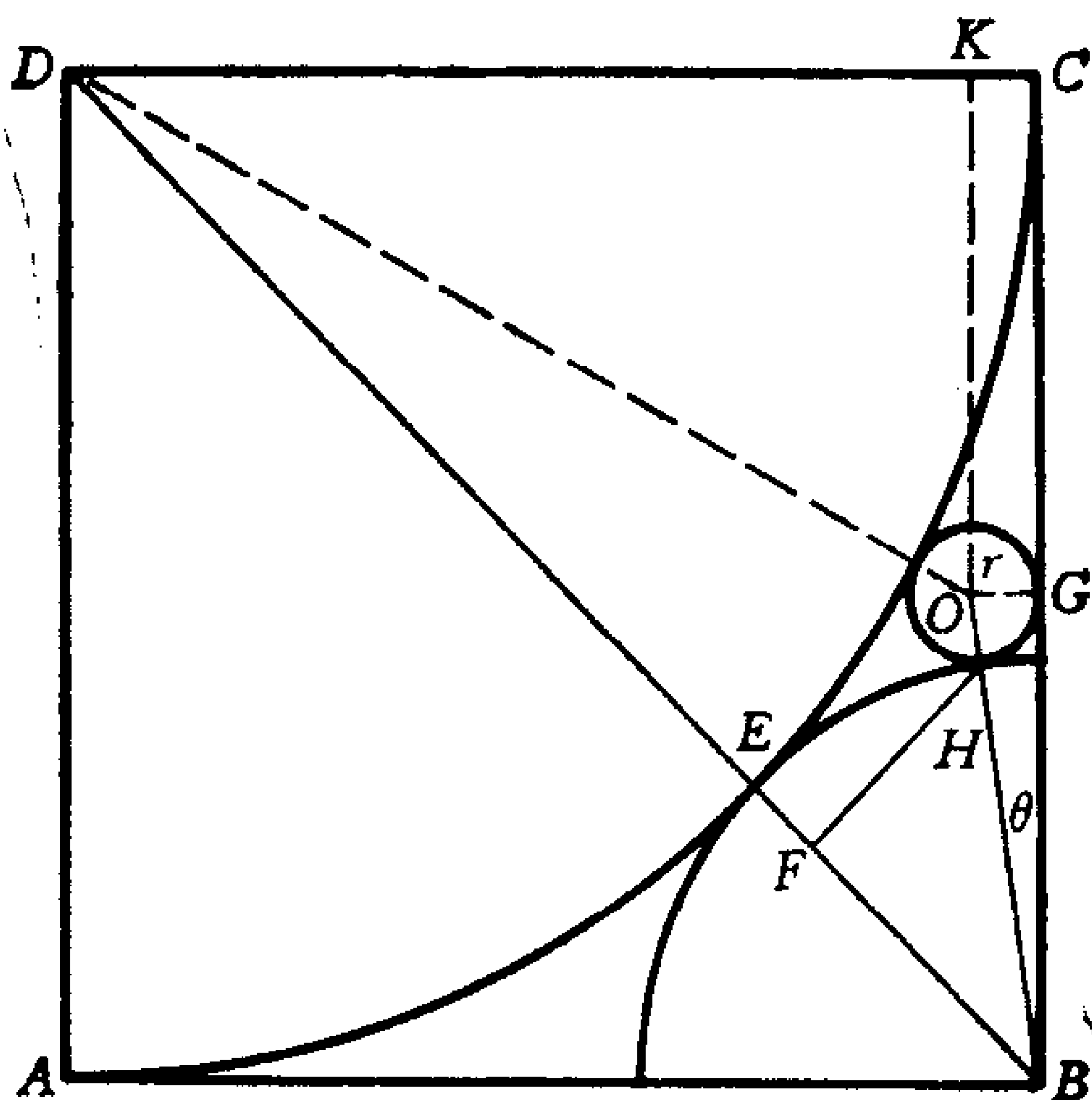


图 9

圆(O)与 BC,(B)及(D)相切，设其半径 $OG=OH=r$。作 $OK\perp CD$, $HF\perp DB$。由△DKO 得，

$$CG=KO=\sqrt{(OD)^2-(DK)^2}=\sqrt{(1+r)^2-(1-r)^2}$$
$$=2\sqrt{r}.$$

由△GOB得，$GB=\sqrt{(OB)^2-(OG)^2}=\sqrt{(r+\sqrt{2}-1)^2-r^2}$.

另外，$GB=CB-CG=1-2\sqrt{r}$。从而有，$r(3-\sqrt{2})-2\sqrt{r}+\sqrt{2}-1=0$。解得 $\sqrt{r}=(2\sqrt{2}-1)/7$。于是 $\tan\theta=OG/GB=1/7$. $\tan\angle HBF=\tan(\pi/4-\theta)=3/4$,因此△HFB是3:4:5三角形([4])。

在一条长度等于 $x^2+y^2=z^2$ 的直线段 AB 上，作一个半圆，并作垂线 CD，垂足是 D，且 $AD=x^2$, $DB=y^2$。则 $CD=\sqrt{x^2y^2}=xy$, $AC=xz$, $CB=yz$。因此，图10中直角三角形 ADC, CDB 及 ACB 的边长之比等于 $x:y:z$。

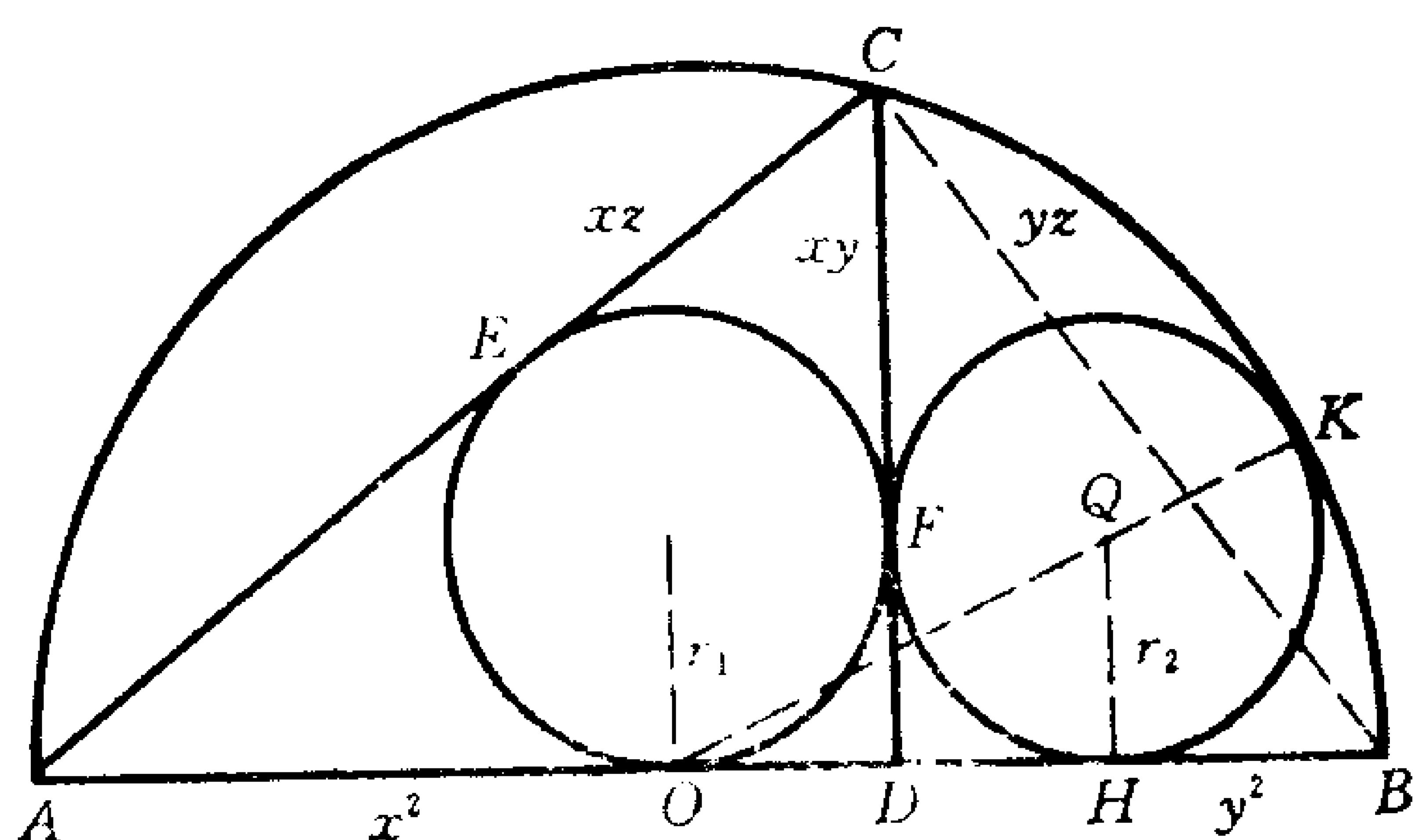


图 10

因为 $CE=CF, AE=AO$ 及 $DO=DF=r_1$，所以 $\triangle ACD$ 的内切圆半径 $r_1=x(x+y-z)/2$。

由 BD, DC 及弧 CB 围成的曲边三角形的内切圆半径 r_2 是直角三角形 OHQ 的一条直角边。因此有

$$r_2^2+(AD+r_2-AB/2)^2=(OK-r_2)^2,$$

即

$$r_2^2+(x^2-z^2/2+r_2)^2=(z^2/2-r_2)^2,$$

它的正根是 $r_2=x(z-x)$。

如果 $r_1=r_2$，则 $z-x=y/3$。因为 $(z+x)(z-x)=y^2$，所以$3z=5y, 4z=5x$。

因此 $y:x:z=3:4:5$。

5. Steiner 链

在图11中，1/4圆(A)和(B)的半径是 2，半圆(O_1)的半径是 1. 在由(A)，(B)和(O_1)所围的曲边三角形中，(O_2)是最大的内切圆，沿着顶点B的尖角往下连续得到 Steiner 圆

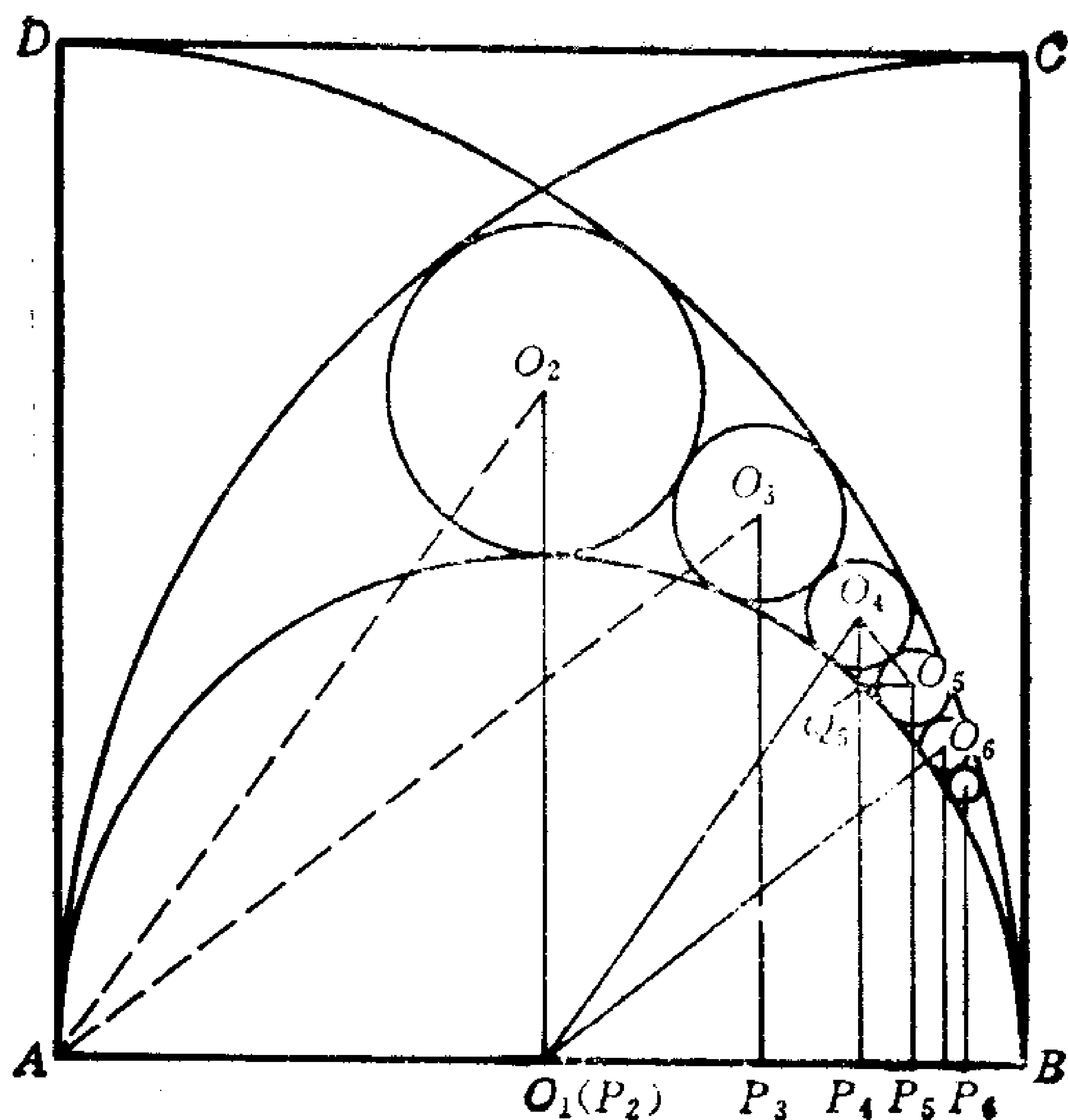


图 11

链(O_n)，$n>1$。从各个圆心 O_n 引 AB的垂线，垂足记为 P_n。则 O_1 与 P_2 重合。可以证得(O_n)的半径 $r_n=2/(n^2+2)$。另外，直角三角形 $O_1O_nP_n$ 的边长是

$$O_1O_n=(n^2+4)/(n^2+2),$$

$$O_nP_n=4n/(n^2+2),$$

$$O_1P_n=(n^2-4)/(n^2+2).$$

因此三角形 $O_1O_4P_4$ 及 $O_1O_6P_6$ 都是3:4:5三角形。

可以推得 $AP_n=1+O_1P_n=(2n^2-2)/(n^2+2)$及 $AO_n=(2n^2+2)/(n^2+2)$。因此三角形 AP_2O_2 及 AP_3O_3 都是 3:4:5 三角形。

从 O_n 作 $O_{n-1}P_{n-1}$的垂线，设垂足是 Q_n，则

$$O_nQ_n = P_nO_1 - P_{n-1}O_1$$
$$= 6(2n-1)/[(n^2+2)(n^2-2n+3)],$$
$$O_{n-1}Q_n = O_{n-1}P_{n-1} - O_nP_n$$
$$= 4(n+1)(n-2)/[(n^2+2)(n^2-2n+3)],$$
而且圆心连线
$$O_{n-1}O_n = r_{n-1} + r_n$$
$$= 2(2n^2-2n+5)/[(n^2+2)(n^2-2n+3)].$$
从而$\triangle O_4O_5Q_5$是3:4:5三角形。

6. 由折叠而得的网格

用折叠或作图的方法，将正方形的顶点与其对边的中点连起来，得到交叠的两个四角星，如图12。这种样式的图形包含有24个3:4:5三角形，如$\triangle BFA$，$\triangle GFE$和$\triangle CDE$那样类型的各有 8 个([6])。

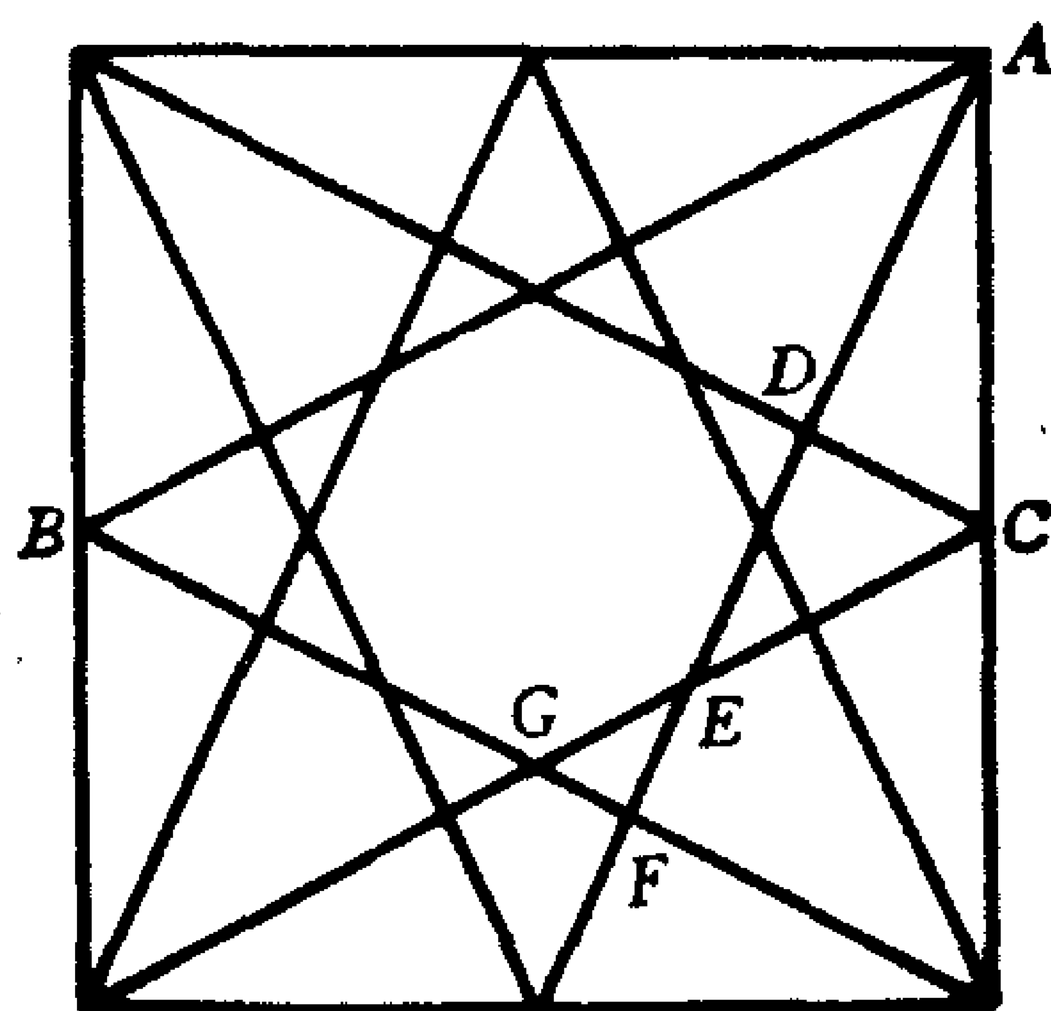


图 12

7. 其他地方出现的3:4:5三角形

如果用直线切割一个三角形成四块，使这四块又能拼在

一起，组成与原三角形相似的两个三角形，则又会出现3:4:5三角形。在任意一个三角形 ABC 的边 AB,BC,CA 上分别取点 D,E,F，使 $AD/AB=CE/CB=AF/AC=1/5$。设 G 是 AC 上一点使 $AF=FG$，又设 H 是 DE 的中点。则图13中的线段 DE,DF，及 GH 表示本节开头时说的3个切割。所需的一个三角形是 BDE，另一个三角形由先将 $\triangle ADF$ 绕 F 旋转直至点 A 与 G 重合，再绕 H 旋转梯形 $GCEH$ 直至点 D 和 E 重合而得到([7],[8])。这三个三角形的相应边长之比是

$$(2DF+EC):BE:BC=3:4:5。$$

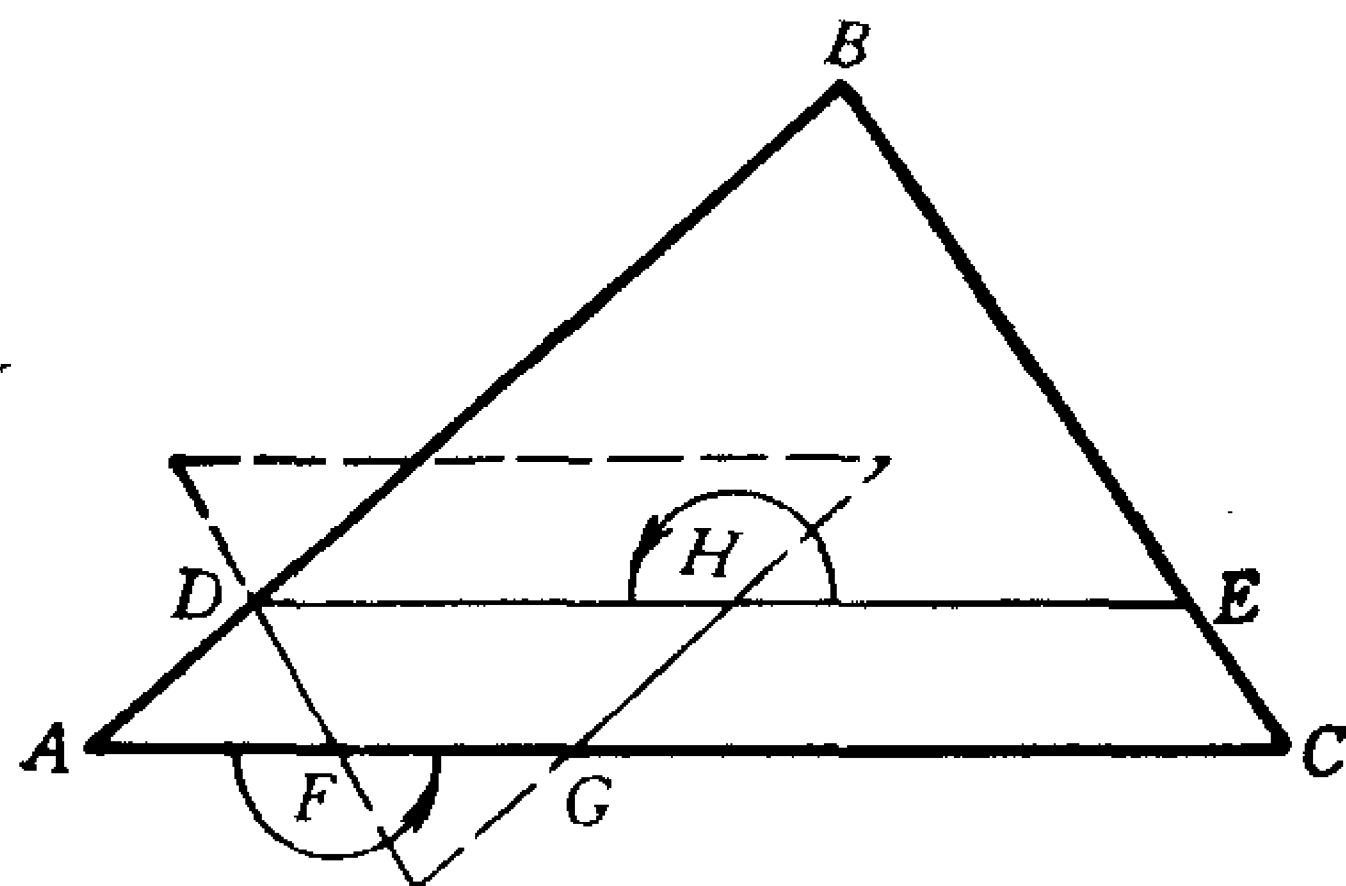


图 13

如果延长正方形 $ABCD$ 的边 CD 到 E 使 $CD=DE$，并连结 BE 交 AD 于 F，连结 CA 交 BF 于 G，见图14，则正

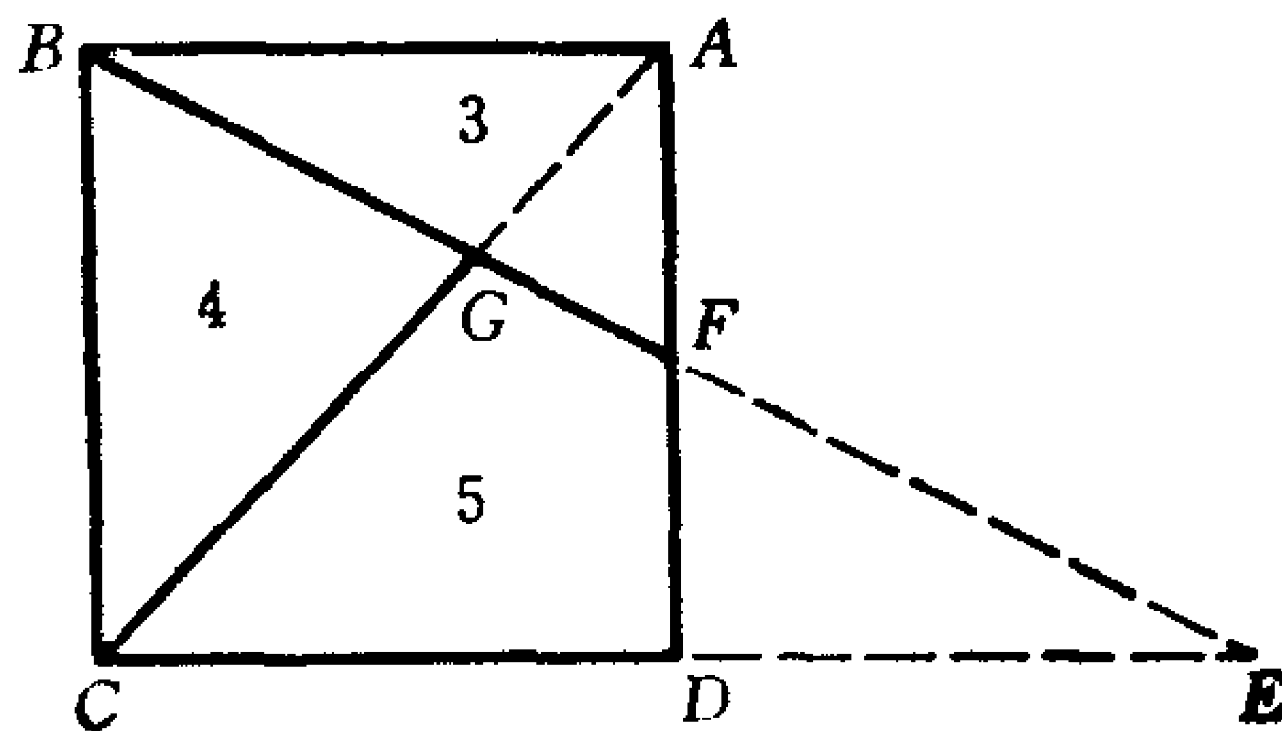


图 14

方形 $ABCD$ 被分割，且 $S_{\triangle ABF}:S_{\triangle BGC}:S_{四边形GCDF}=3:4:5$（[9]）。

机灵的读者会发现，在许多其它的图形中都会含有3:4:5三角形。对于能引起他们注意的任意这种发现，作者都会感到很高兴。

参考文献

[1] C. W. Trigg and M. W. Fleck, E432, *Amer. Math. Monthly*, 48 (1941), 267—268.

[2] John Casey, A Sequel to Euclid, Longmans, Green, and Co., London, 1884, p. 118, Prop. 7.

[3] R. A. Johnson, Modern Geometry, Houghton Mifflin, Boston, 1929, p. 113.

[4] Leon Bankoff and W. J. Cherry, A 3:4:5 right triangle, *Amer. Math. Monthly*, 61 (1954), 473—474.

[5] Leon Bankoff and W. J. Blundon, An infinite sequence of Pythagorean triangles, *Amer. Math. Monthly*, 62 (1955), 734—736.

[6] C. W. Trigg, Configuration generated by folding a square, *Scripta Math.*, 21 (1955), 77—80.

[7] Michael Goldberg and Aaron Buchman, A dissection of a triangle, *Amer. Math. Monthly*, 58 (1951), 112.

[8] C. W. Trigg, Mathematical Quickies, McGraw-Hill, New York, 1967, p. 140.

[9] Leon Bankoff and D. M. Brown, A square dissection, *Math. Magazin*, 29 (1955), 110—112.

（朱学贤译，刘　勇校）

三角形内心和旁心的重心坐标①

三角形有许多令人感兴趣的特殊点，例如内心（内切圆圆心）和旁心（旁切圆圆心）。本文给出它们的重心坐标的极其简单的表达式，这在合成构形性质的经典讨论中有用。

重心坐标是几何学中的一个重要概念，见[2]。

先介绍一维的重心坐标。设 A_1,A_2 是2个固定点，选取从 A_1 到 A_2 的方向为正向。从而正方向上的线段（例如 A_1A_2）的长度（不妨仍记为 A_1A_2）为正，而 $A_2A_1<0$。设 $t_1+t_2\neq0$。将质量 t_1 与 t_2（解释为电量可能更合适些，因为 t_1 或 t_2 可以是负数，但沿袭习惯术语）分别置于点 A_1 和 A_2（这样的点也被称为"加权点"），则可以确定唯一的重心 P，见图1。

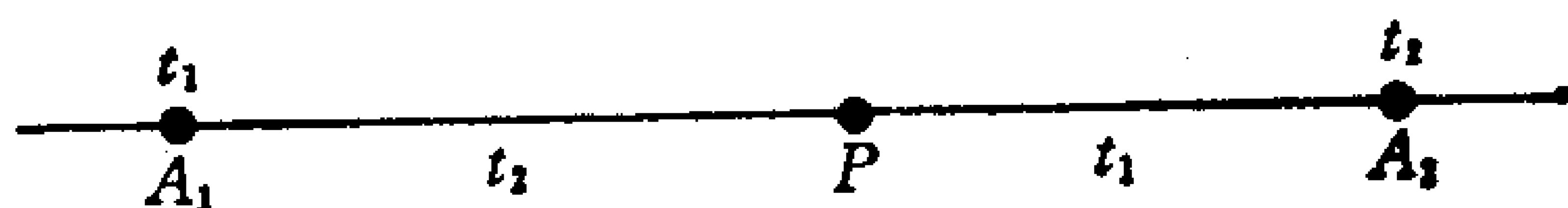


图 1

下列事实是显见的：

(i) 若 $t_1=0$ 则 $P=A_2$，若 $t_2=0$ 则 $P=A_1$。

(ii) 若 t_1 与 t_2 均为正或均为负，则 P 点在线段 A_1A_2 上。

(iii) 若 $t_1>-t_2>0$，则 P 点在线段 A_2A_1 的延长线上；

① 编译自文末的参考文献[1]和[2]。

若 $t_2 > -t_1 > 0$，则 P 点在线段 A_1A_2 的延长线上。

反过来，对于直线 A_1A_2 上的任意一点 P，可以求得 t_1，t_2，使

$$\frac{t_2}{t_1}=\frac{A_1P}{PA_2} \text{ 或 } \frac{t_1}{t_2}=\frac{PA_2}{A_1P},$$

即，P 是将质量 t_1 与 t_2 分别置于 A_1 和 A_2 后的重心。称 (t_1, t_2) 为 P 的重心坐标。因为将质量 μt_1 与 $\mu t_2 (\mu \neq 0)$ 分别置于 A_1 和 A_2 后的重心仍为 P，所以重心坐标满足齐次条件

$$(t_1, t_2)=(\mu t_1, \mu t_2), \quad \mu \neq 0.$$

类似地，可以建立平面上关于给定的三角形 ABC（称为"坐标三角形"）的重心坐标（Möbius 在1827年就看到了）。规定按逆时针方向排列3个顶点的三角形的面积为正，否则为负。设 $t_1+t_2+t_3 \neq 0$。将质量 t_1, t_2 与 t_3 分别置于三角形的顶点 A, B 和 C，则可以确定唯一的重心 P（注意：P 点不一定在 $\triangle ABC$ 内），记其坐标为 (t_1, t_2, t_3)。特别地，$(1,0,0)$，$(0,1,0)$ 和 $(0,0,1)$ 分别是点 A, B, C；$(1,1,1)$ 是 $\triangle ABC$ 的形心；$(0, t_2, t_3)$ 是直线 BC 上的点，其关于 B 和 C 的一维重心坐标是 (t_2, t_3)（注意：方向 $A \to B, B \to C$ 及 $C \to A$ 是正向），等等。反过来，对于 $\triangle ABC$ 所在平面上的任意一点 P，可以求得 (t_1, t_2, t_3)，使 P 是将质量 t_1, t_2 和 t_3 分别置于 A, B 和 C 后的重心。例如，见图2，设连结 A 与 P 的直线交 BC 于 Q，先求出 Q 关于 B 和 C 的一维重心坐标 (t_2, t_3)，然后设 P 是将质量 t_1 和 (t_2+t_3) 分别置于 A 和 Q 的重心，求出 t_1。定义 (t_1, t_2, t_3) 为 P 点关于坐标三角形 ABC 的重心坐标。与一维的情形相同，它也满足齐次条件：

$$(t_1, t_2, t_3)=(\mu t_1, \mu t_2, \mu t_3), \quad \mu \neq 0.$$

（条件

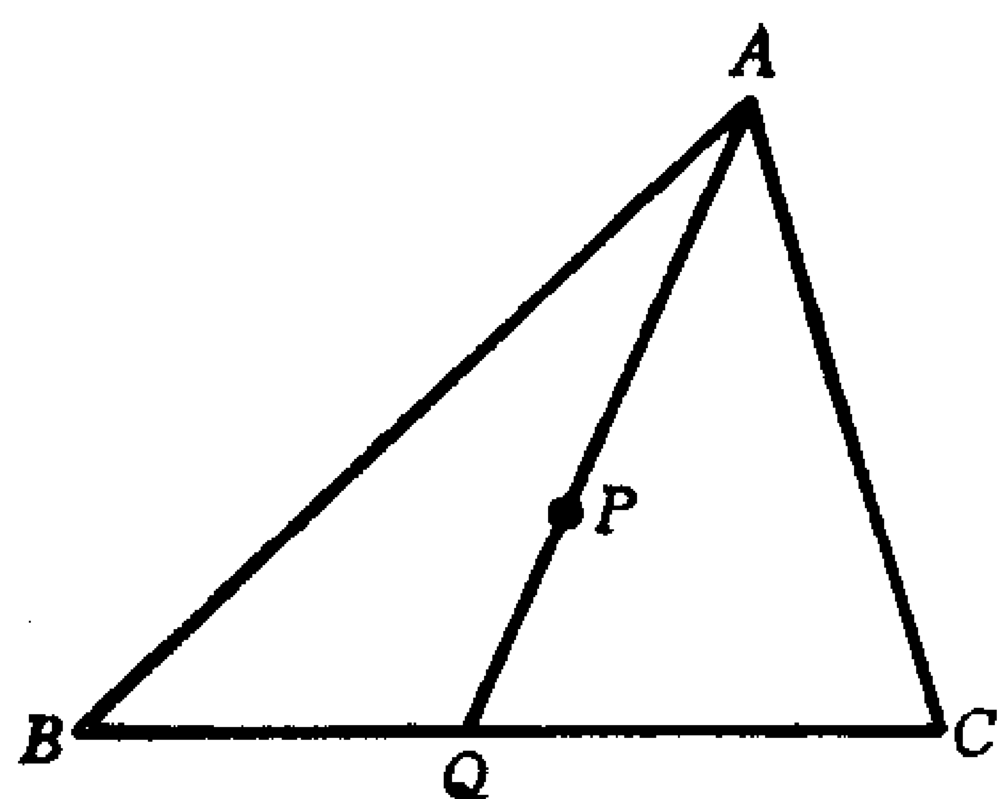


图 2

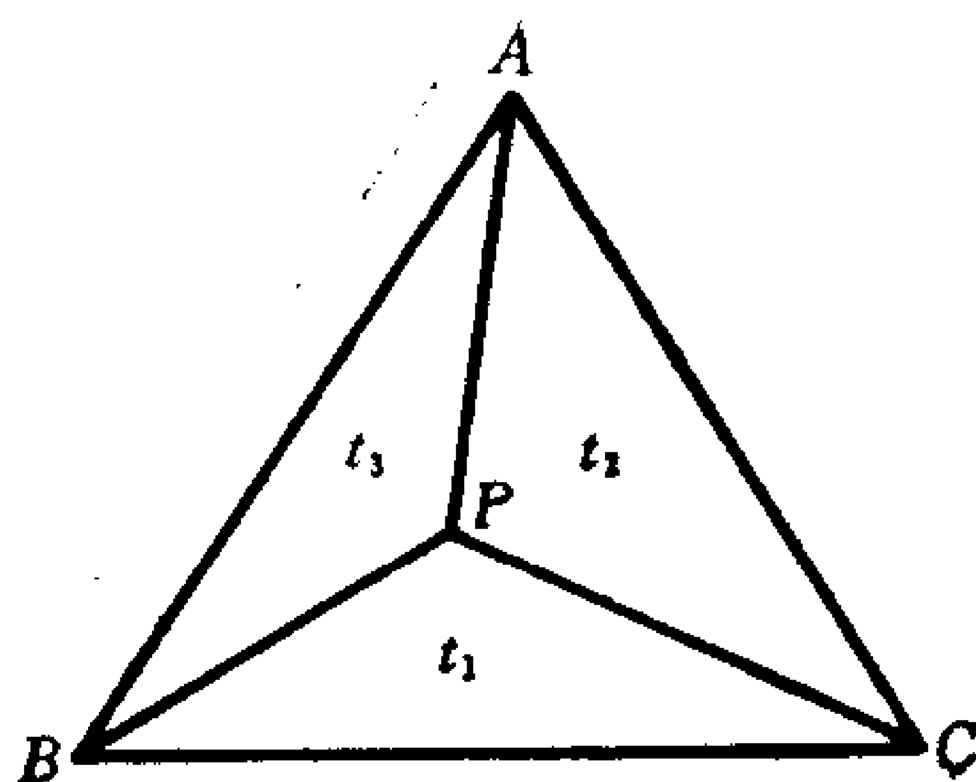


图 3

$$t_1+t_2+t_3\neq 0,$$

可以使我们规范化重心坐标，即使

$$t_1+t_2+t_3=1.$$

这种规范化的重心坐标称为面积坐标，它的意义可从下面的命题1看到.)

连结 PA,PB 及 PC，得3个三角形 $\triangle PBC$，$\triangle PCA$ 及 $\triangle PAB$（注意顶点排列的次序）。

命题1（[2]） 设 P 点关于坐标三角形 ABC 的重心坐标为(t_1,t_2,t_3)，则有

$$S_{\triangle PBC}:S_{\triangle PCA}:S_{\triangle PAB}=t_1:t_2:t_3 \qquad (1)$$

（其中，若某项为零，则按通常的意义理解）。

证明 先设 P 在$\triangle ABC$ 内，见图3。则有，例如

$$\frac{t_3}{t_2}=\frac{BQ}{QC}=\frac{S_{\triangle ABQ}}{S_{\triangle AQC}}=\frac{S_{\triangle PBQ}}{S_{\triangle PQC}}$$

$$=\frac{S_{\triangle ABQ}-S_{\triangle PBQ}}{S_{\triangle AQC}-S_{\triangle PQC}}=\frac{S_{\triangle PAB}}{S_{\triangle PCA}}.$$

从而易证(1)式成立；若 P 在$\triangle ABC$ 外，只要注意到关于三角形面积正负的规定，同样可以证得(1)式；若 P 在 $\triangle ABC$

的某条边上，则某个 $t_i=0$，且易证(1)式成立。

命题 1 证得。

现在建立三角形的内心和旁心的重心坐标表达式。不妨设$\triangle ABC$ 即为坐标三角形，A,B,C 所对的边（及其边长）分别记为 a,b,c。记内心为 I，3 个旁心分别为 I_a,I_b 和 I_c（足标表示与该旁切圆相切的边）。见图 4。

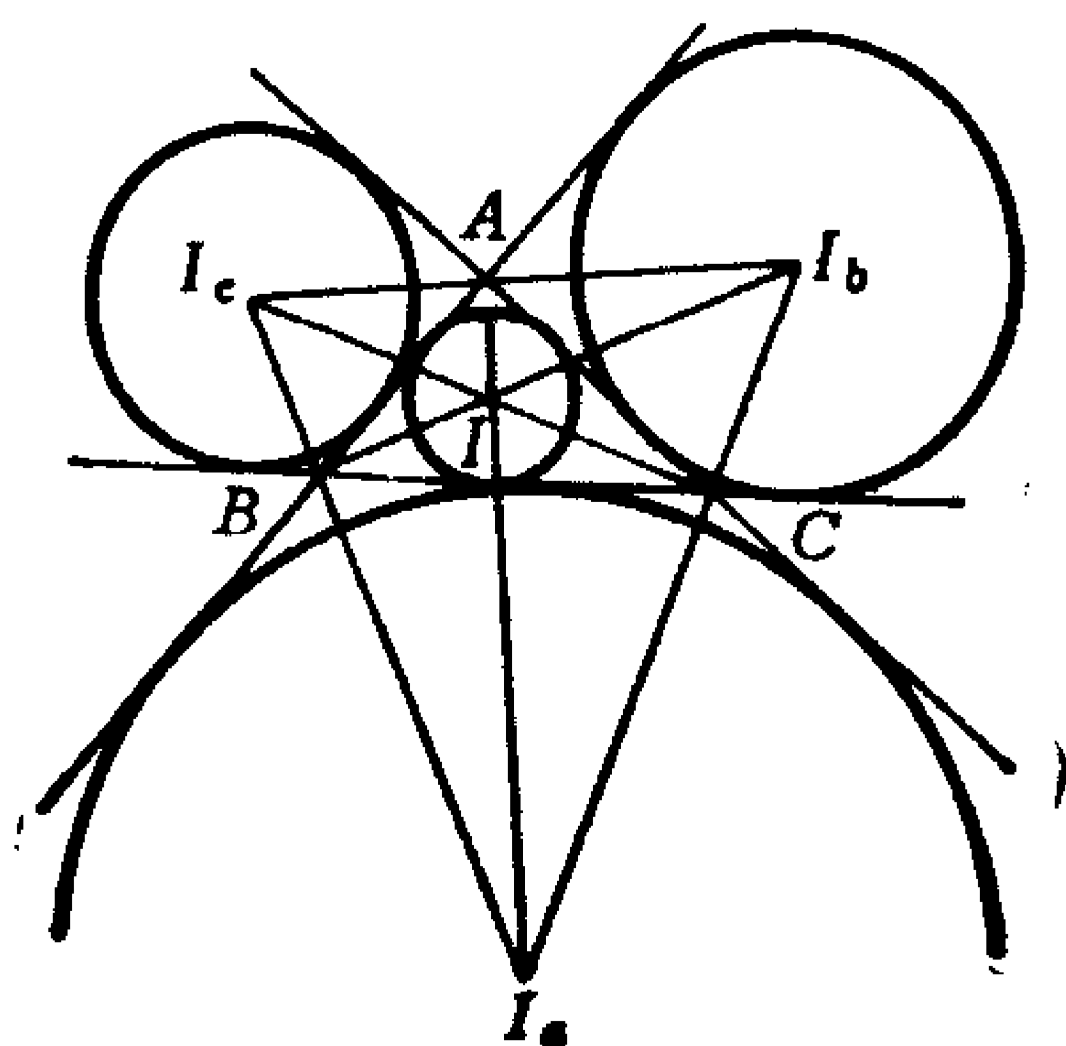


图 4

命题2([1]) 点 I,I_a,I_b 和 I_c 可唯一地表成

$$I=\frac{aA+bB+cC}{a+b+c},\quad I_a=\frac{-aA+bB+cC}{-a+b+c},$$

$$I_b=\frac{aA-bB+cC}{a+b+c},\quad I_c=\frac{aA+bB-cC}{a+b-c},$$

其中的点都取重心坐标形式。

证明 设内切圆半径为 r，则有

$$S_{\triangle IBC}=\frac{1}{2}ar,\quad S_{\triangle ICA}=\frac{1}{2}br,\quad S_{\triangle IAB}=\frac{1}{2}cr.\tag{2}$$

设内心 I 的重心坐标为(t_1,t_2,t_3)。由(1)和(2)得

$$t_1:t_2:t_3=S_{\triangle IBC}:S_{\triangle ICA}:S_{\triangle IAB}$$
$$=a:b:c,$$

从而证得内心 I 的重心坐标表达式。类似可得旁心的表达式，不过要注意面积的正负。

命题 2 得证。

可以将重心坐标的概念推广到 n 维空间。从而，例如命题 2 的结论及证明能推广到 n 维单形的内心和旁心，只是要用顶点所对的$(n-1)$维单形的容量代替所对边的边长。

参考文献

[1] P. Tondeur, Barycentric representation for the incenter and excenters of a triangle, *Amer. Math. Monthly*, 94 (1987), 975—976.

[2] H. S. M. Coxeter, Introduction to Geometry, John Wiley & Sons, INC., 2-nd, 1989. (Chap. 13)

(朱学贤编译，潘承彪校)

用几何变换证明Euclid几何定理①

Ross L.Finney

观察平面的变换，能够得到一些Euclid几何定理的漂亮的证明，这些定理很容易叙述，而且其中有许多能够从随手画出的图形中看出来。定理的证明既不用坐标系也不用向量，其中最简单的证明只要对旋转和平移比较熟悉就够了。我们先从某几个定理出发，然后引进相似变换，把文献中一系列孤立的结果看成若干一般性定理的特殊情形。在此给出的证明的附带结论是：所得到的关于四边形的结果，并不要求四边形是凸四边形或简单四边形（例如可以比较定理1，定理3与[5]中的对应定理）。

引理1 如果等腰三角形ZMX和YMW在点M处有直角，那么$\overline{YX}$和$\overline{ZW}$垂直且相等（参看图1）。

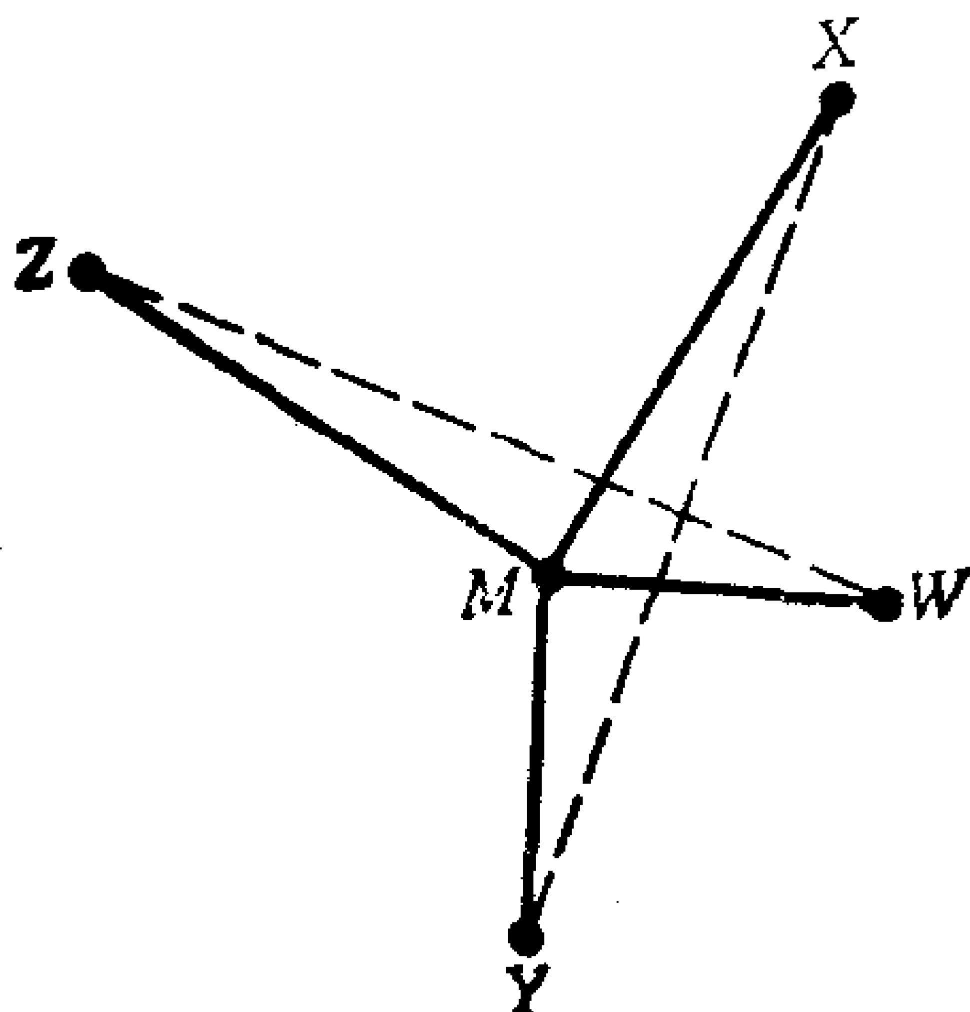


图 1

① Dynamic proofs of Euclidean theorems, *Math. Magazine*, 9月—10月（1970），177—185。文后的附录由陈维桓编写。

证明 这是因为绕M作逆时针的90°旋转，即M_{90}，把Y转到W，把X转到Z。

引理2 如果Z和X是在三角形ABC两边上向其外部所作的两个正方形的中心，点M是三角形第三边的中点，那么ZMX是等腰三角形，并且角M为直角（参看图2）。

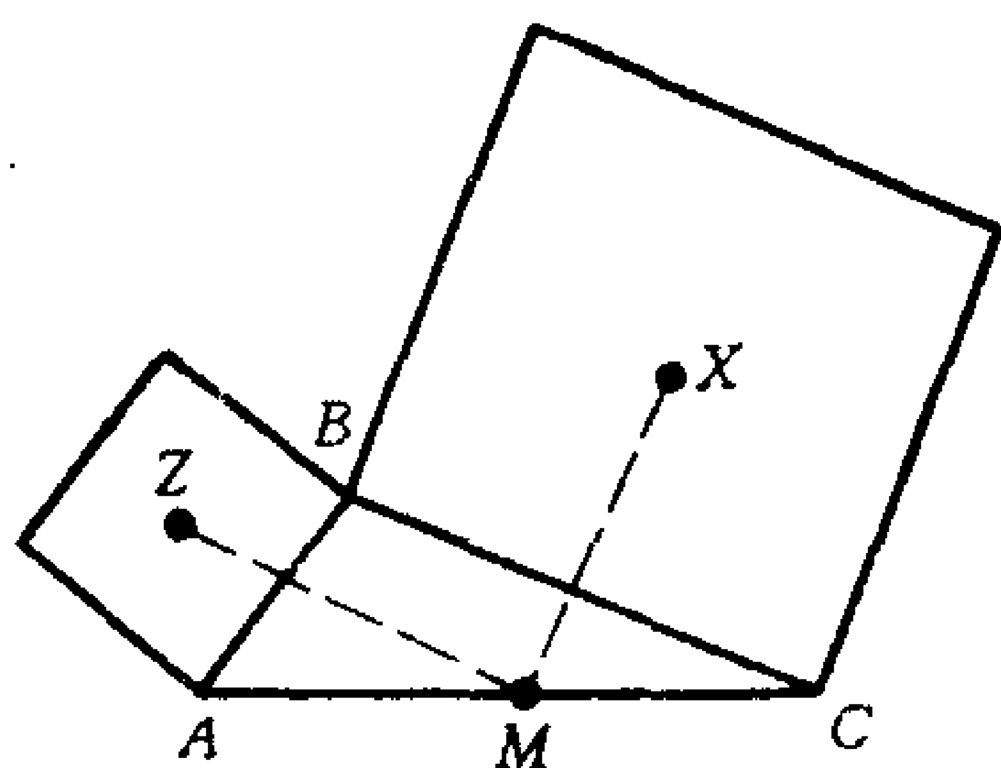


图 2

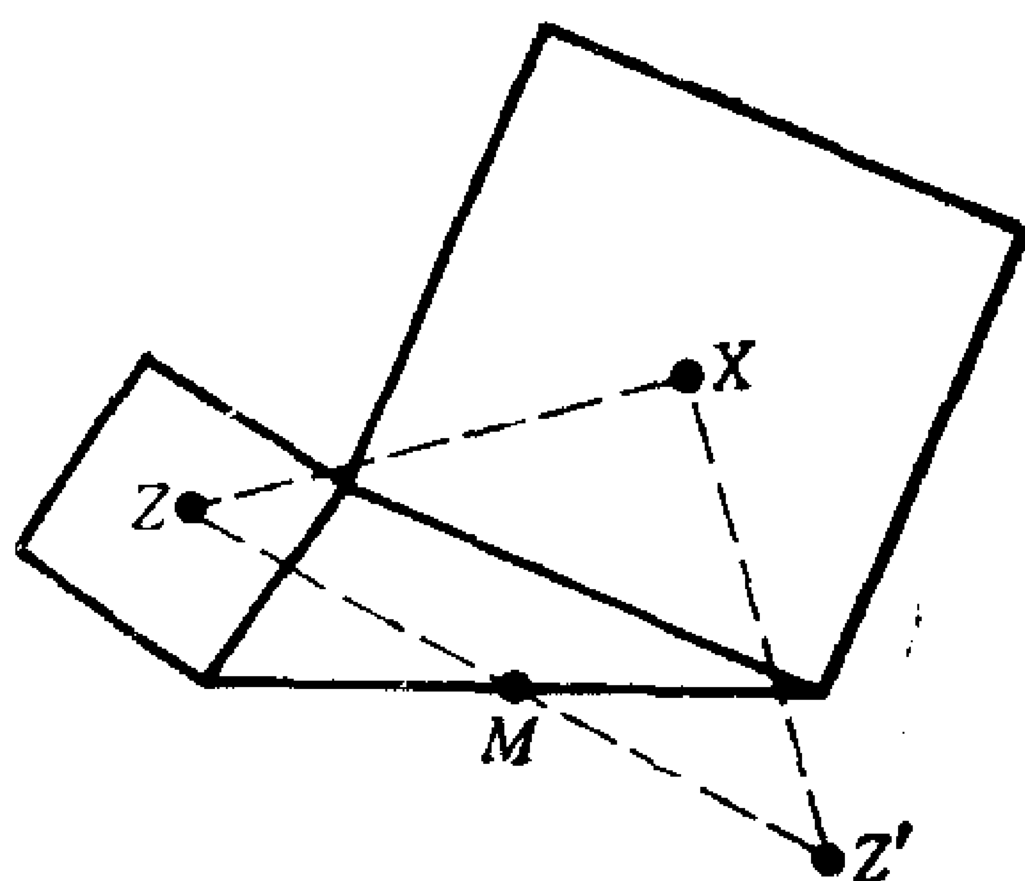


图 3

证明 首先我们注意到：平面上围绕两个不同点的旋转的合成是平面上的一个平移（此时，两个旋转的旋转角之和是360°的整数倍），或者是围绕第三点的一个旋转，其旋转角恰好是原先两个旋转的旋转角之和(参看下面的附录)．现考虑图2中围绕点Z,X,M各个旋转的合成$T=M_{180}X_{90}Z_{90}$，因为各个旋转角之和是360°的整数倍，故T是平移（或恒同变换）．在这个合成运动下，A先到达B，然后到达C，最后回到A．所以A是T的不动点，因而T只能是恒同变换（具有一个不动点的平移只能是恒同变换）。点Z在上述运动下逐次的映象是：

$$Z_{90}(Z)=Z,$$

$$X_{90}(Z)=\text{某点}Z',$$

$$M_{180}(Z') = T(Z) = Z.$$

现在得到的图形如图3所示.由于$X_{90}(Z) = Z'$,故$ZX = XZ'$,且三角形ZXZ'在X处有直角。因为$M_{180}(Z') = Z$,得知M是ZZ'的中点。现在M是等腰直角三角形ZXZ'的斜边的中点,因此ZMX是等腰三角形,且在点M处有直角。

如果在三角形两条边上向内作正方形,而不是向外作正方形,则引理2的结论仍成立。在证明中,仅需用顺时针旋转Z_{-90}和X_{-90}替代逆时针Z_{90}和X_{90}即可。

定理1 ([1,3]) 如果X,Y和Z,W分别是在四边形的两对对边上向外作正方形的中心,那么$\overline{YX}$和$\overline{WZ}$垂直且相等。

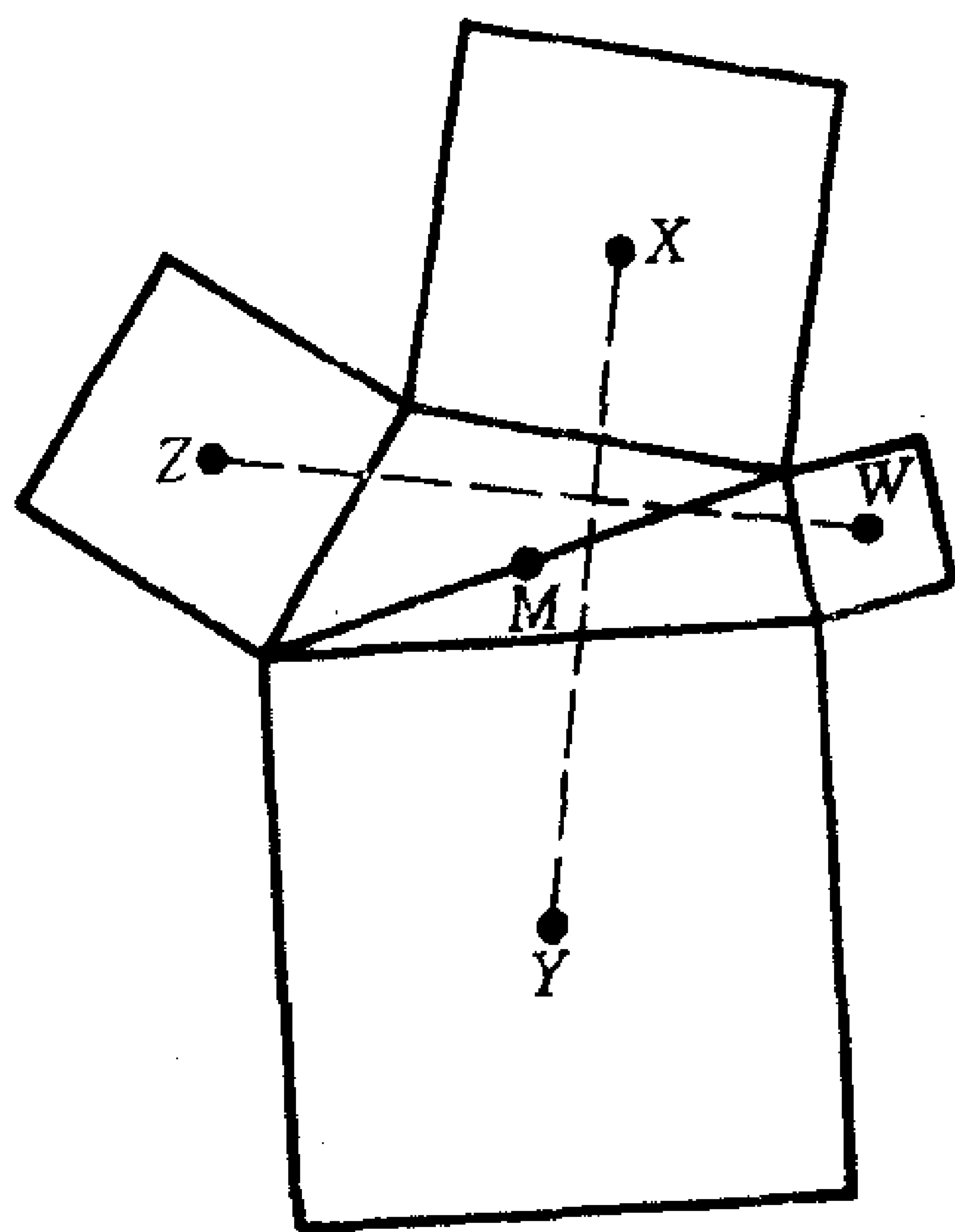


图 4

证明 利用引理2,命M是四边形一条对角线的中点,并考虑旋转M_{90}即可。

如果这些正方形都向四边形的内部,则定理1仍然成立。

证明相同，M_{90}把其中一条线段变为另一条线段。

如果该四边形恰好是平行四边形，那么$ZXWY$成为一个正方形，因为四个三角形ZMX，XMW，WMY和YMZ全都是等腰三角形，并且在M处为直角。

应该强调，已知四边形不必是凸四边形，也不必假定是简单四边形。当四边形不具有明显的内部时，为了放置这些正方形，只要在两种可能的方向中取定一种方向绕四边形行进，然后把正方形放在行进方向的右边（参看图5）。

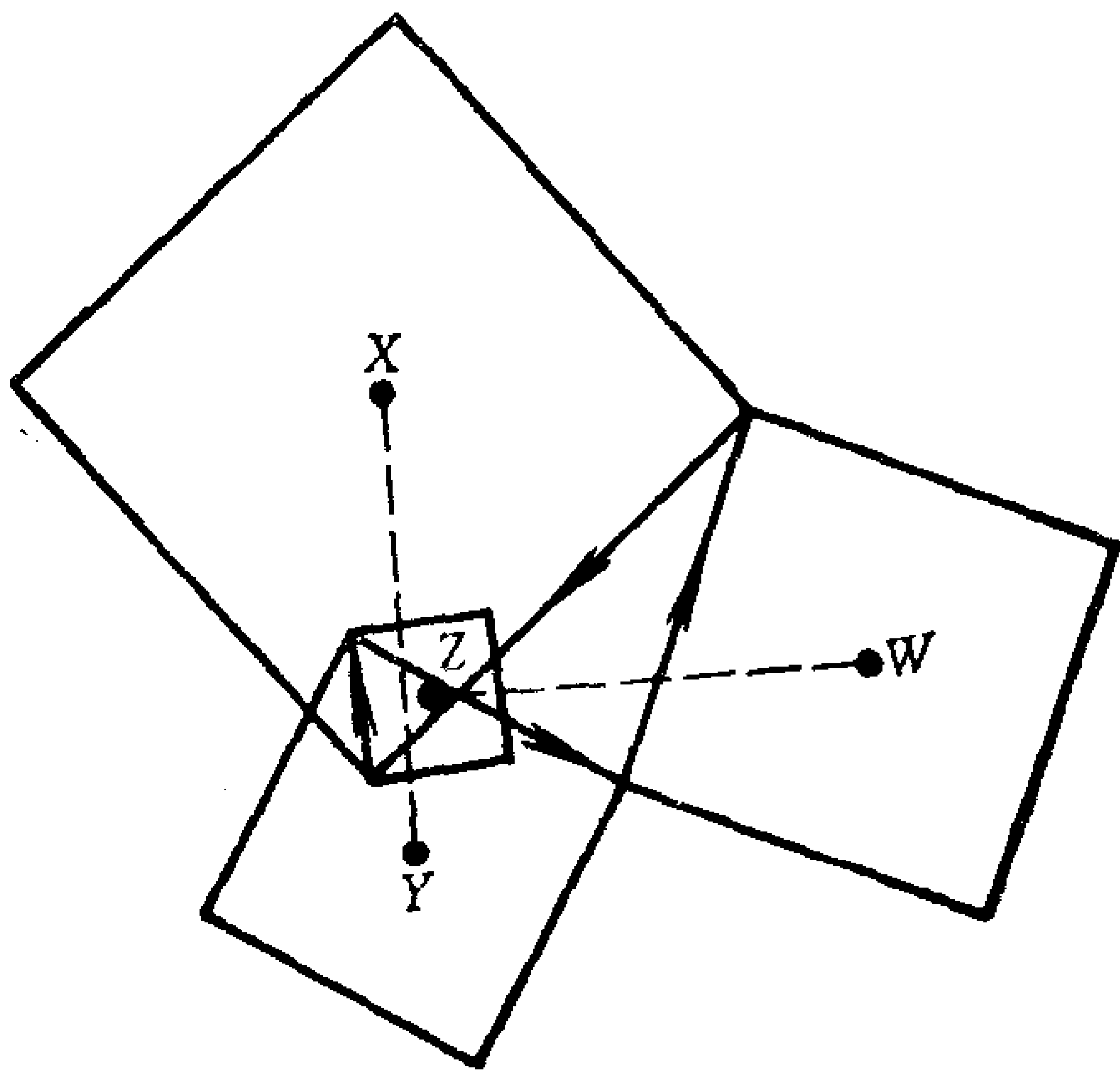


图 5

若把四边形的一边收缩为一点，我们便得到关于三角形上的正方形的定理。

定理2([1,3]) 如果在三角形各边上向其外部作正方形，则连接其中两个正方形中心的线段垂直且等于连接第三个正方形中心与它所对的三角形顶点的线段。

证明 参看图6，并用如同引理1中的旋转 M_{90}即可。

定理2有如下的推论：连接正方形中心与其所对的三角形顶点的三条直线共点（图7）。原因是这些连线恰好是三角形XYZ的高线。

Euclid 几何中有如下的一个定理：四边形各边的中点是一个平行四边形的顶点．此外，还可证明：

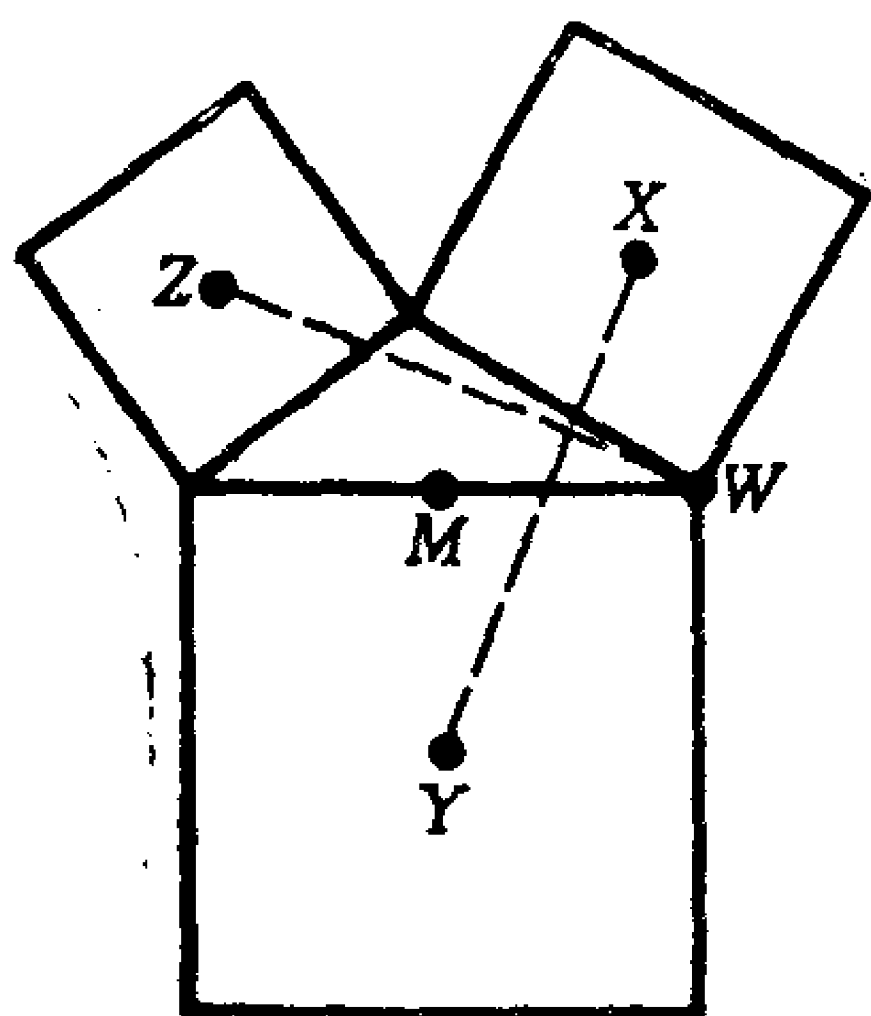


图 6

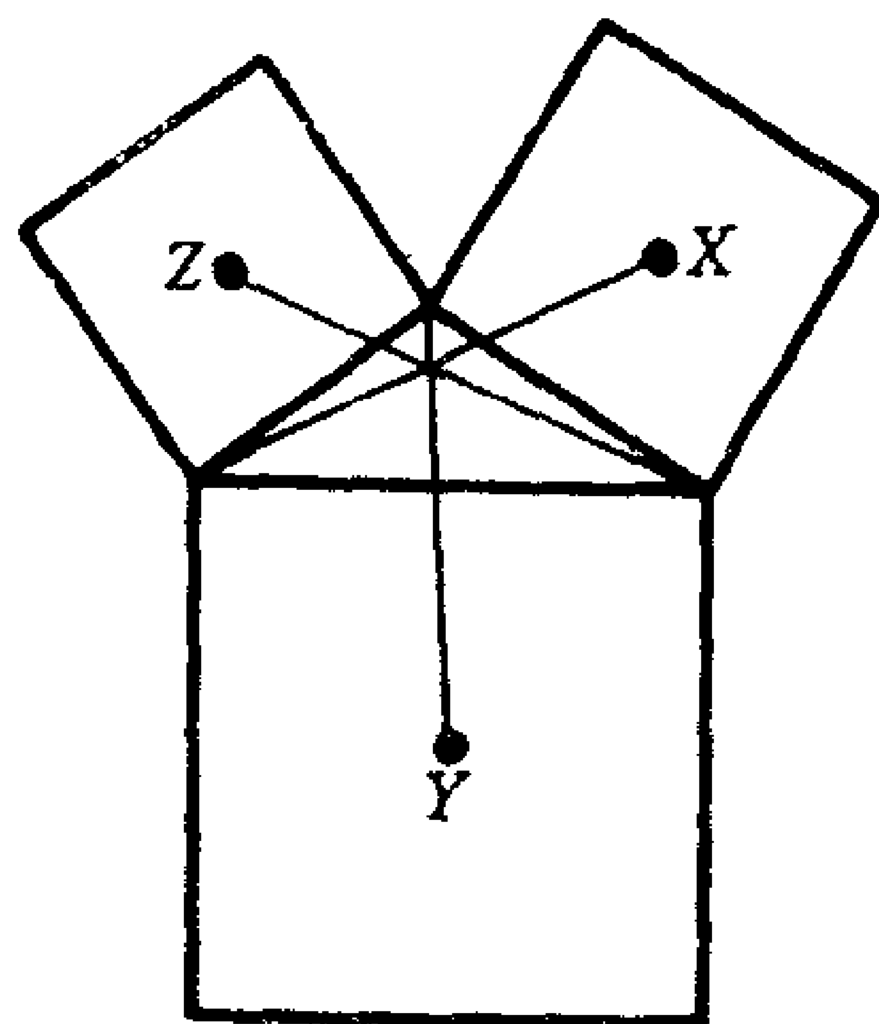


图 7

定理3 在四边形各边上交替向四边形内部和外部所作的等边三角形的顶点Z,X,W,Y是一个平行四边形的顶点（参看图8）。

证明 合成变换$C_{-60}A_{60}$是一个平移，它把Y变到W，把Z变到X。于是$\overline{YZ}$和$\overline{WX}$平行且相等。

前面提到的Euclid几何关于中点的定理和定理3是一族定理中的两个特例。

定理4（[2]） 如果Z,X,W,Y是在四边形各边上适当放置的四个相似三角形的顶点（参看图9），那么$ZXWY$是平行四边形。

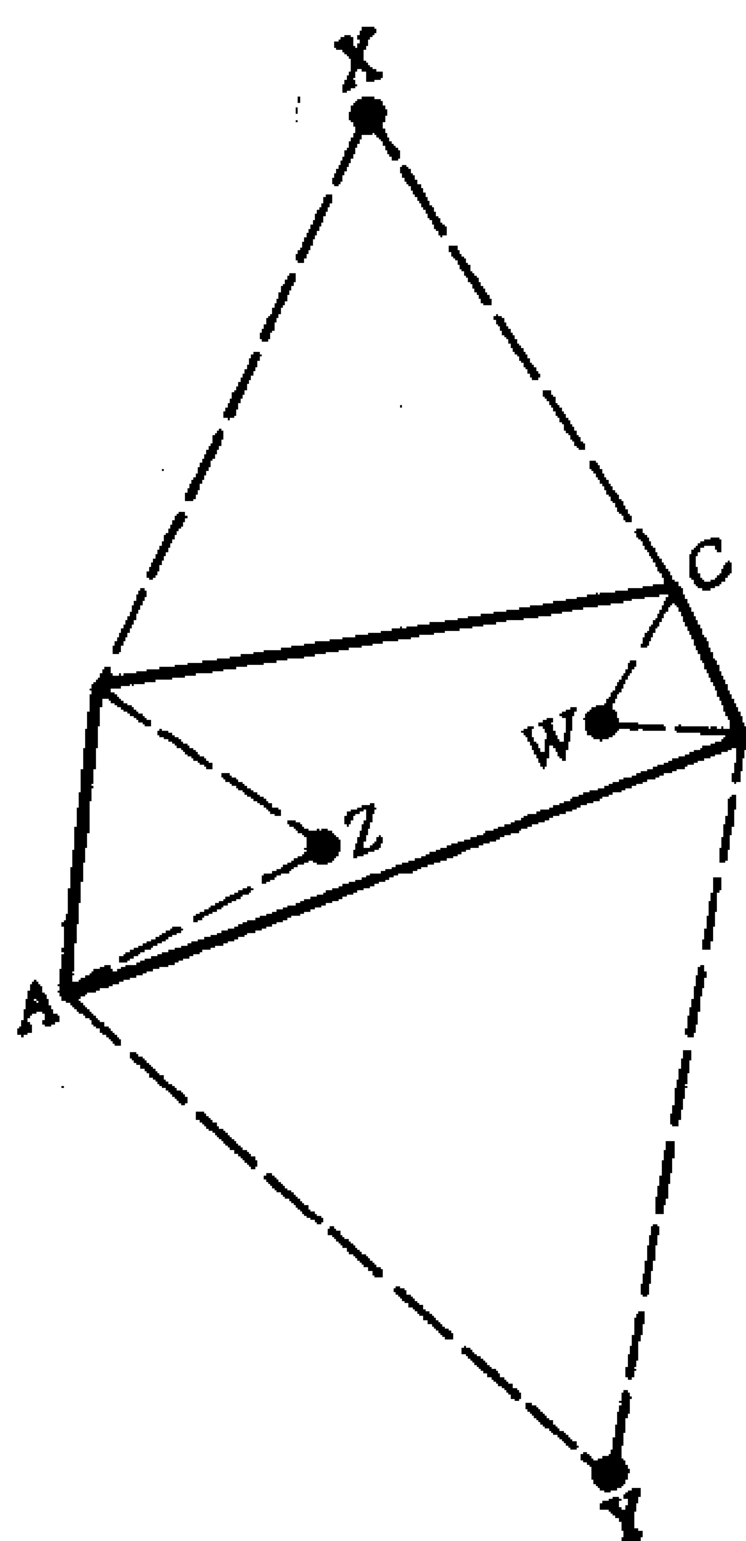


图 8

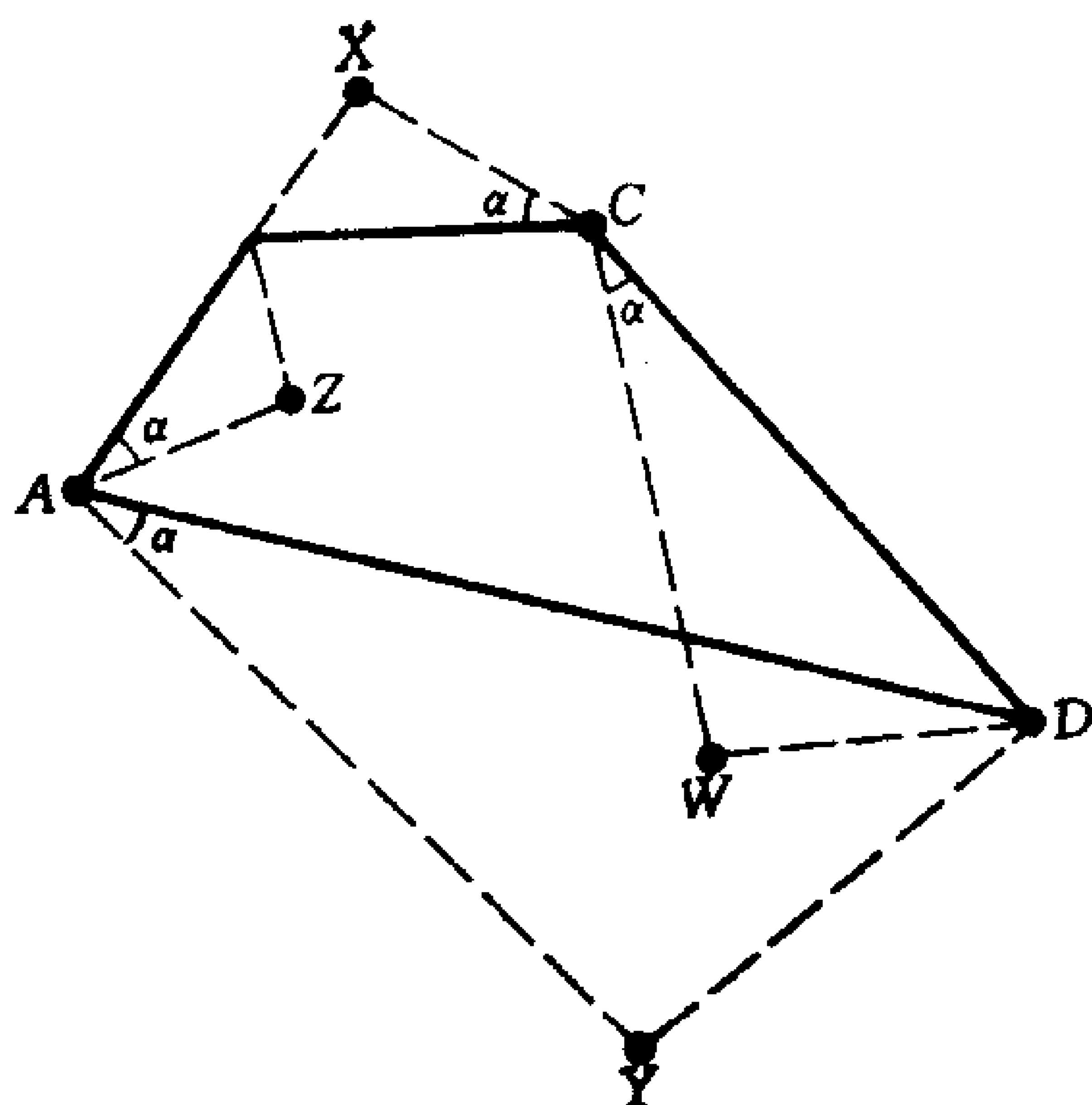


图 9

证明 设α是四个相似三角形在点A和点C处的内角，且设r为AY与AD之比。设$A^{1/r}$表示以A为中心、比例系数为$1/r$的中心相似变换。设C^r表示以C为中心、比例系数为r的中心相似变换。一般说来，以两个不同点为中心的相似变换的合成是以第三点为中心的相似变换，其比例系数为原先两个中心相似变换的比例系数的乘积；特别是，若两个中心相似变换的比例系数的乘积为1，则它们的合成是一个平移。于是合成变换$C_{-\alpha}C^r A^{1/r} A_\alpha$是一个平移(参看下面的附录)，它把$Y$变到$W$，把$Z$变到$X$。

如果$\alpha=0$，则得到关于中点的定理的推广。如果$r=1$，$\alpha=60°$，我们便有定理3。又，该四边形不需要假定是凸四边形或简单四边形。将一条边收缩为一点时，便能得到几个

关于三角形的很好的定理。

利用相似变换，我们能够证明关于三角形的一个定理（见定理5），它有确实是出人意料的推论。其中三个推论如下所述：

推论1 假设在任意一个三角形的两边向外作内角为30°,60°,90°的三角形，如图10所示。令Z和X表示这两个三角形的外顶点，令M是已知三角形第三边的中点。那么ZMX是等边三角形。如果在已知三角形两边上向内作内角为30°,60°,90°的三角形，则ZMX仍是等边三角形。

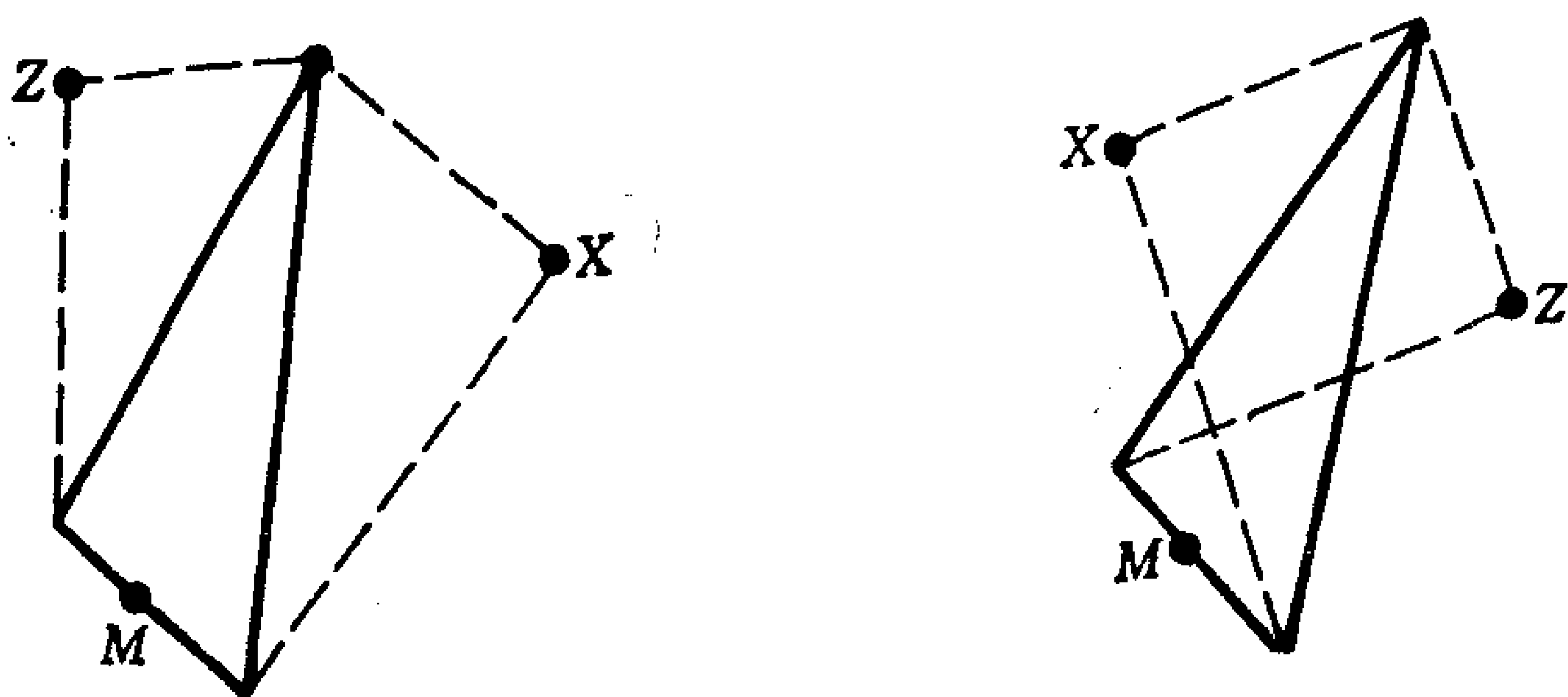


图 10

推论2（Napoleon定理） 在任一三角形的各边向外部作等边三角形，则其中心X,Z,M是一个等边三角形的顶点（参看图11）。

推论3（[5，练习23]） 假定在任一三角形各边上作三个等边三角形，其中两个向外，一个向内。设M是向内三角形的中心，Z和X是向外的三角形的顶点。那么三角形ZMX是在M点的内角为120°的等腰三角形（见图12）。

为了使定理5的叙述简单一些，我们假定角是“一般的”，

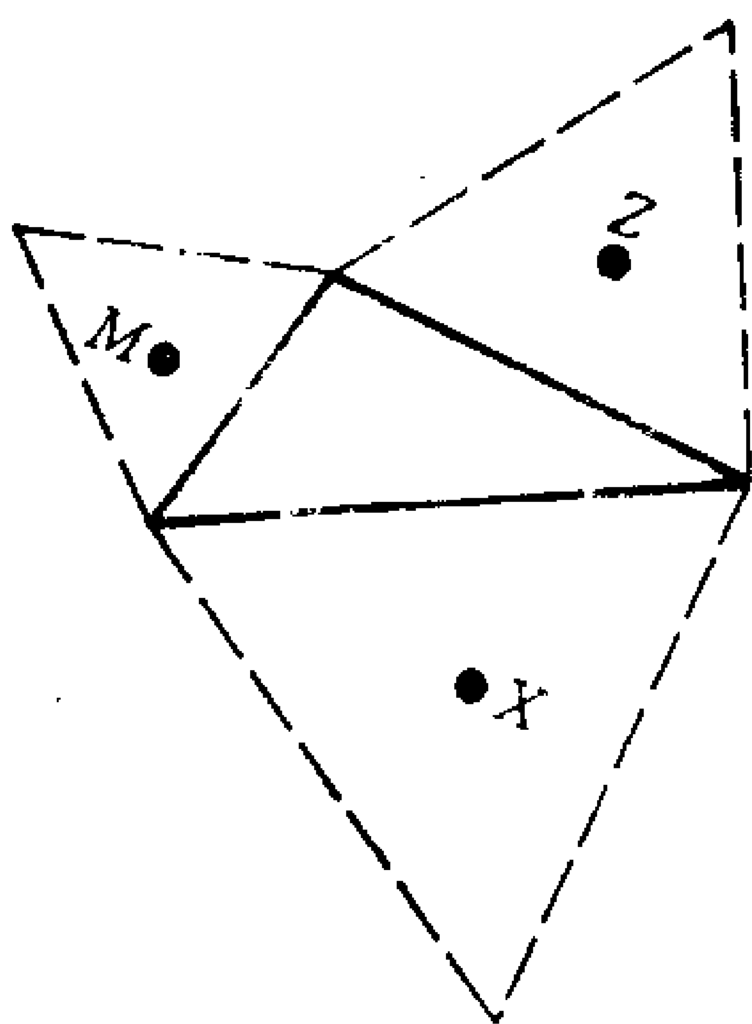


图 11

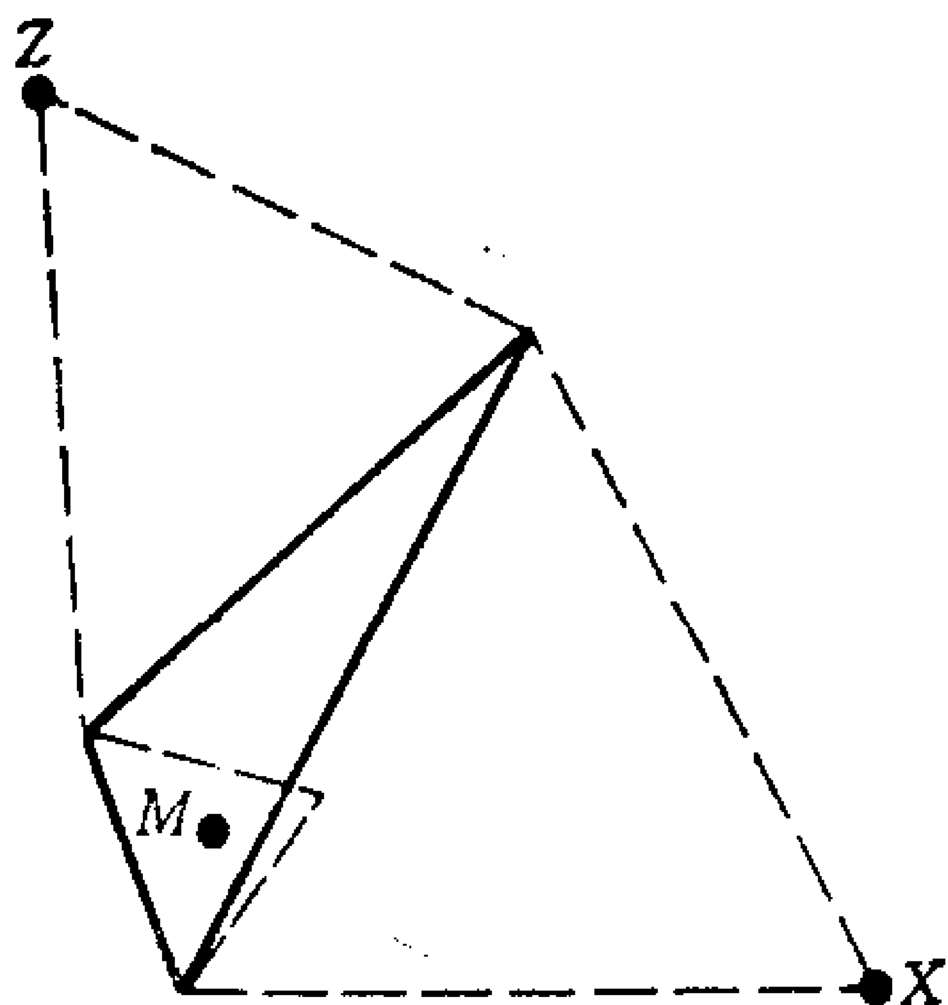


图 12

也就是它们的取值可增减360°的整数倍，且以逆时针方式定向。例如，图13表示了两个角：角 CBA 为60°，420°等等而角 ABC 为300°，-60°等。

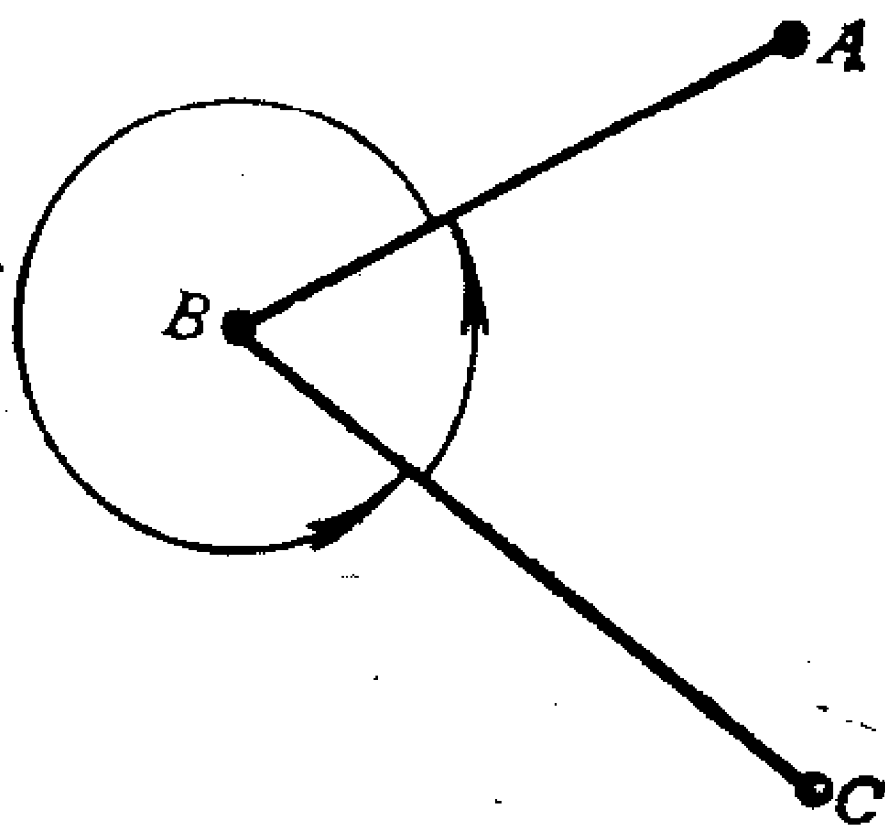


图 13

定理5 设 BZC 和 CXA 是在 $\triangle ABC$ 两边上所作的两个非退化的相似三角形（对应顶点如给定顺序所示），它们都朝三角形 ABC 的外部或内部。设 $\angle BZC$ 和 $\angle CXA$ 的值为 β。设 M 为平面上一点，它到 A 和 B 等距且使 $\angle BMA=2\beta$。则 $MZ=MX$（参看 图14）。

证明 因为三角形BZC和CXA是非退化的，$2\beta\neq 360°$，因而M点确实存在。但是三角形AMB可以是退化的，例如，当$\beta=90°$时，M是$\overline{AB}$的中点。

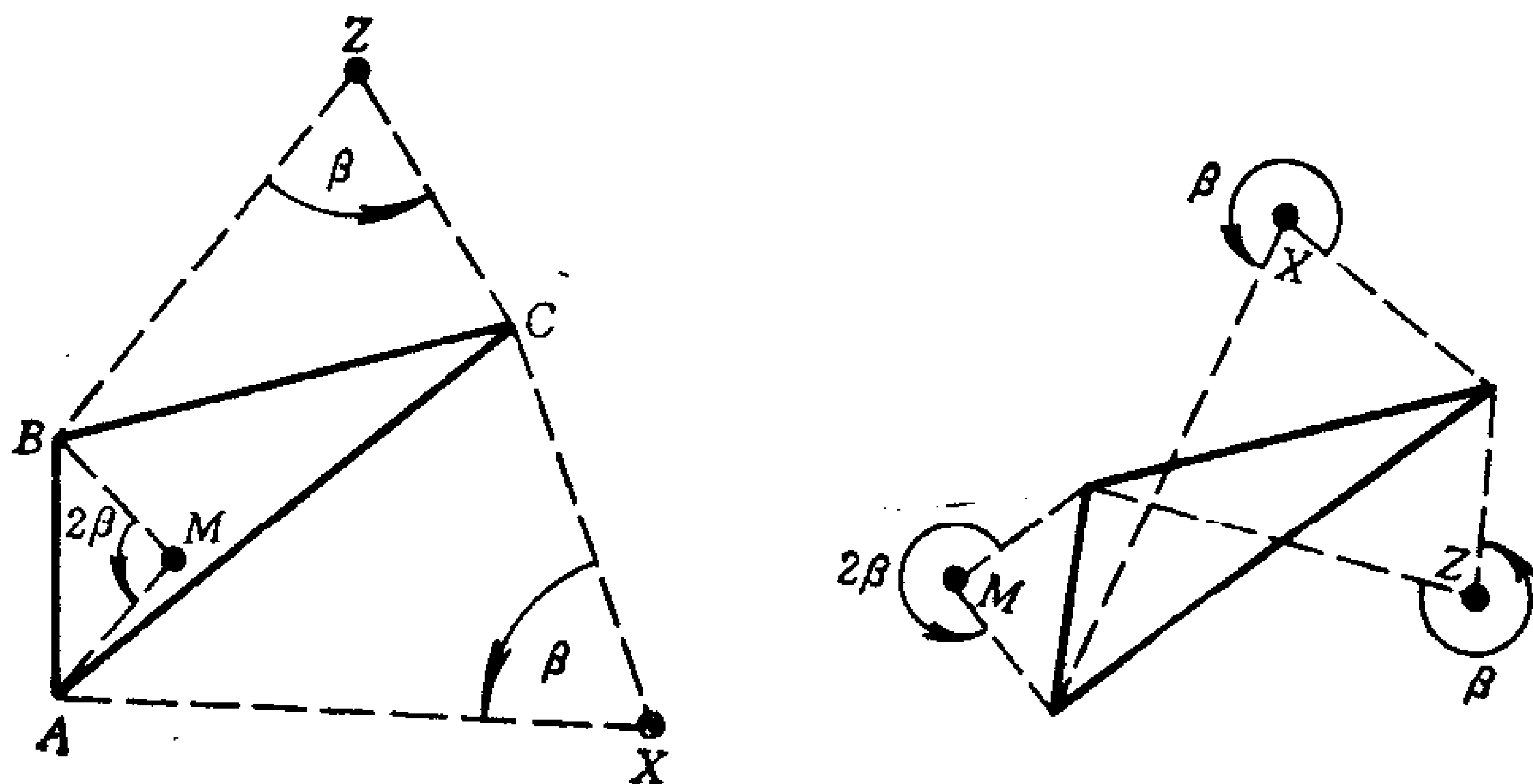


图 14

用r表示比值ZC/BZ。平移$T=X_\beta X^{1/r}Z^r Z_\beta M_{-2\beta}$保持点$A$不变，因而是恒等变换。如果跟踪$M$，便得$M_{-2\beta}(M)=M$，$Z^r Z_\beta(M)$是某点$M'$，于是$X_\beta X^{1/r}(M')=T(M)=M$。因此$MXM'Z$是一个四边形，具有性质$XM'=r\cdot MX$，$ZM'=r\cdot MZ$，并且在$X$和$Z$的角相等。这就是说，三角形$MXM'$和$MZM'$相似。由于它们有一条公共边，因此它们全等，且$MZ=MX$。

推论的证明 如果$\beta=90°$，$r=\sqrt{3}$，那么三角形MXM'和MZM'是内角为30°，60°，90°的三角形。特别地，$MXM'Z$在M的角为60°，因而ZMX是等边三角形，这证明了推论1。

如果$\beta=120°$，$r=1$，则Z，X和M的地位是等同的，不

能加以区分因而 $MZ=MX=XZ$，这就是 Napoleon 定理。

推论3是 $\beta=60°$，$r=1$ 的情况。

如果 $\beta=-60°$，$r=1$，我们有两个在内部和一个在外部的三角形，推论3的结论仍成立。

令 $\beta=-120°$，$r=1$，得到三个向内的三角形，Napoleon 定理的结论仍成立。

引理2也是定理5的推论。如果 $\beta=90°$，$r=1$，正方形就向外。如果 $\beta=-90°$，$r=1$，则正方形就向内。

最后我们给出引理2的推广，它会导致定理1的推广。

定理6([4]) 假设在任意三角形 ABC 边上作两个相似的等腰三角形，它们或者都朝向 ABC 的内部，或者都朝向 ABC 的外部。若令 X 是其中一个三角形的顶点，Z 是另一三角形的垂心，而 M 是三角形 ABC 第三边的中点。令 β 是 X 处的角。则三角形 ZMX 在 M 的角为直角，在 X 的角是 $\beta/2$。

证明是直接的，只要注意到角 AZB 等于 $180°-\beta$。引理

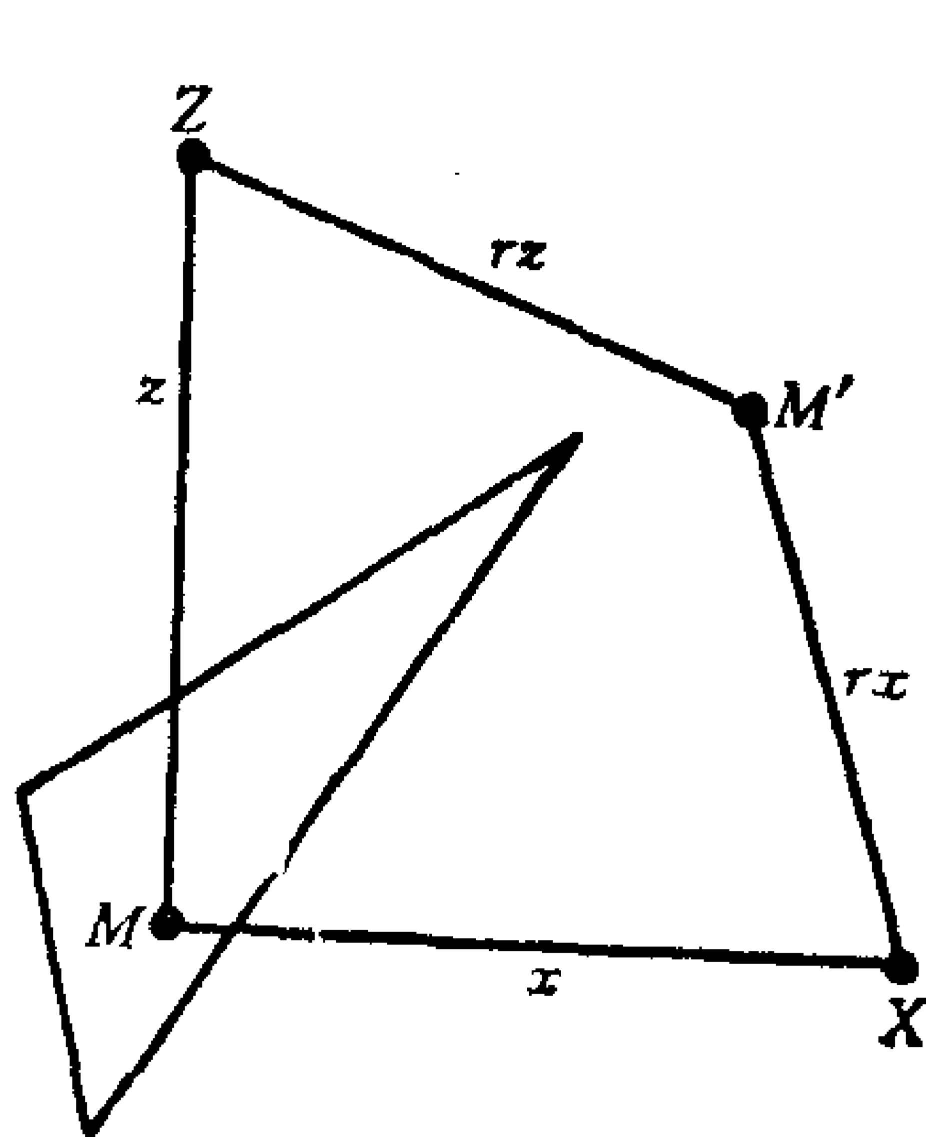


图 15

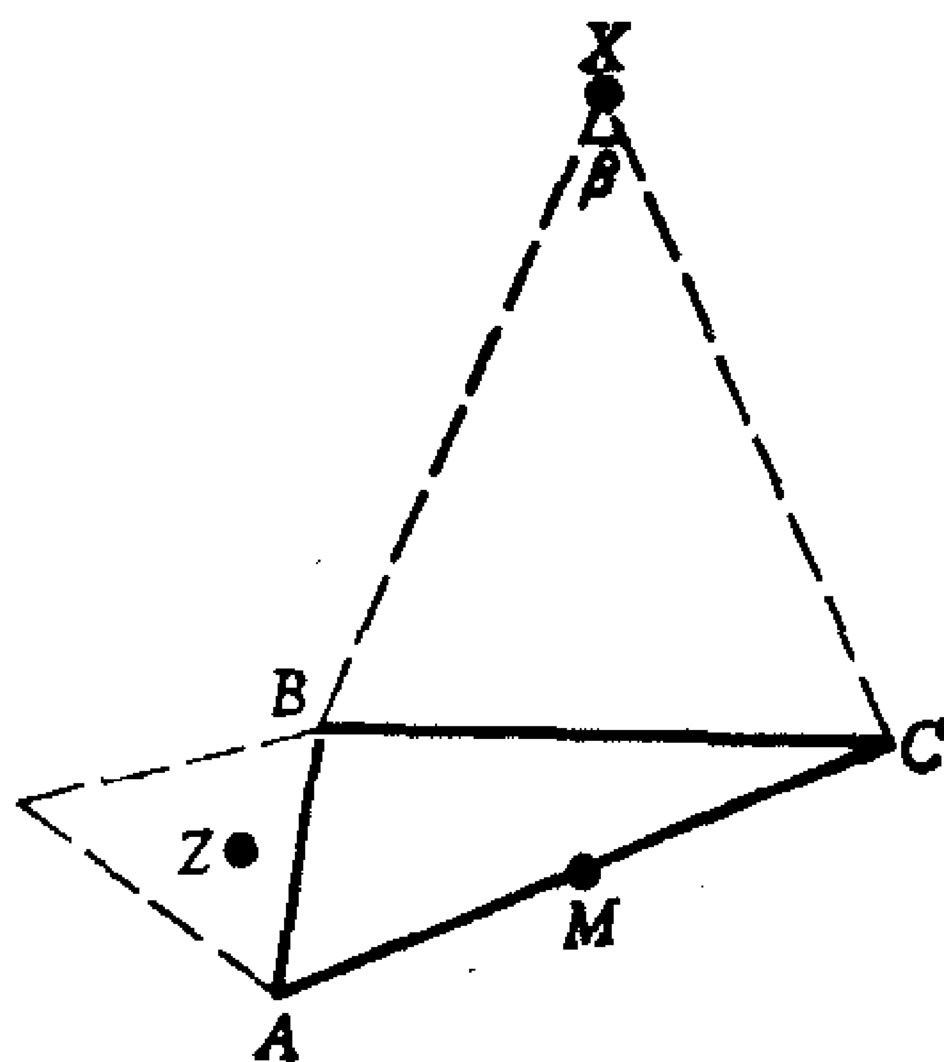


图 16

2是$\beta=90°$的特殊情形。若令h表示X相对于$\overline{BC}$的高，那么$XM/MZ=BC/2h$.在下面的定理7中，我们称后一个数为r.

定理7 假设在四边形的各条边上作相似的等腰三角形，它们都朝向四边形的内部或都朝向四边形的外部。设X和Y是一对相对的三角形的顶点，且设Z和W是另外两个三角形的垂心。那么$\overline{YX}$垂直于$\overline{WZ}$，且$YX/WZ=r$。

证明 令M是对角线$\overline{AC}$的中点。如果这些三角形朝向四边形的外部，那么$M^r M_{90}$把X变到Z，把Y变到W。如果这些三角形朝向四边形的内部，则用$M^{1/r} M_{90}$替代。

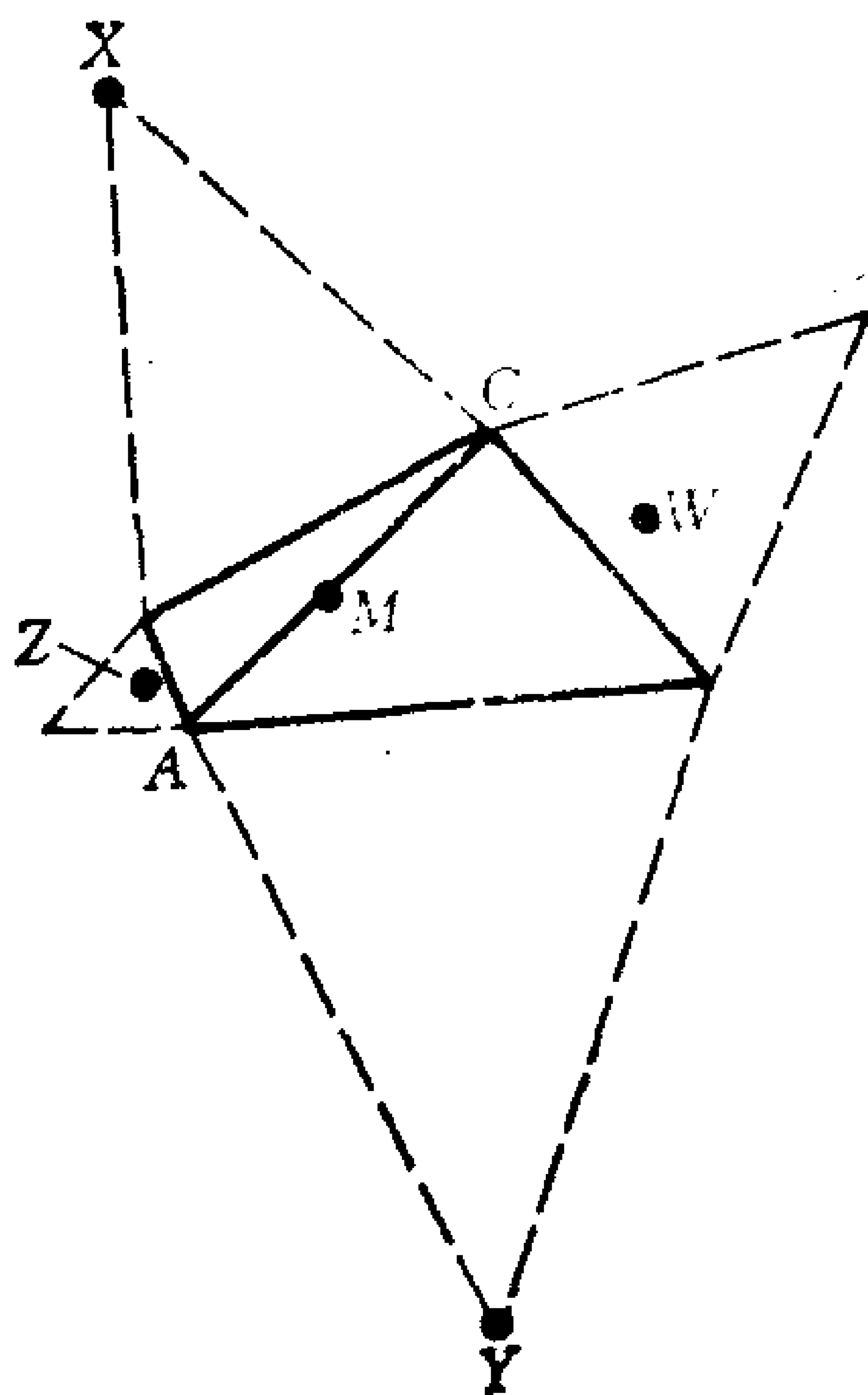


图 17

又，四边形不必是凸的四边形或简单的四边形，并且对于三角形也有类似的定理。

参 考 文 献

[1] M.H.van Aubel, Note concernant les centres des construits sur les côtés d'un polygone guelconque, *Nouv. Corresp. Math.*, 4(1878), 40—44.

[2] ——, Question 56, *Mathesis*, 1(1881), 167.

[3] C.A.Laisant, Sur quelques propriètés des polygones, *Assoc. Franc. Avanc. Sci. Le Havre*, (1877) 142—154.

[4] D.St.J.Jesson, Private communication, 1967.

[5] M.Yaglom, Geometric Transformations, Random House, New York, 1962.

(阮培文编译，陈维桓校)

附　录

按照F.Klein的“爱尔兰根纲令”的观点，Euclid平面几何就是研究平面上的图形在相似变换群及其子群（特别是等距变换群）的作用下的不变性质的学科。平面上的相似变换可以分解为平移、旋转、中心相似变换与反射的合成。关于一条直线的反射必定改变平面上的有向角的符号。我们把保持角的符号不变的相似变换称为正相似变换。平移、旋转、中心相似变换都是正相似变换。为方便起见，我们把关于同一点的中心相似变换与旋转的合成称为关于该点的旋转相似变换。平移、旋转是保持两点之间的距离不变的，因此它们都是正等距变换。非恒同变换的平移没有不动点，旋转角$\neq k\cdot 360°$的旋转有只有一个不动点（旋转中心），比例系数$\neq 1$的中心相似变换也只有一个不动点（相似中心）。有两个不动点的正相似变换必是恒同变换。

我们有以下事实：

(1) 对于平面上两个正全等图形，必有平面上的一个平移或者一个旋转，把其中一个图形变为另一个图形。

证明　平面上两个正全等图形可以用两条长度相等的有向线段作为代表。如果这两条有向线段平行且同向，则在平面上有一个平移将这两条线段合同；若不然，则这两条有向线段的起点连线及终点连线的垂直平分线必相交，以该交点为中心的旋转可以使这两条有向线段合同（见图18）。

(2) 平移与旋转（旋转角$\neq k\cdot 360°$）的合成是旋转。

证明　平移和旋转的合成仍然是正等距变换，所以有向线段$\overrightarrow{AB}$在合成运动作用下得到的是长度相等的有向线段

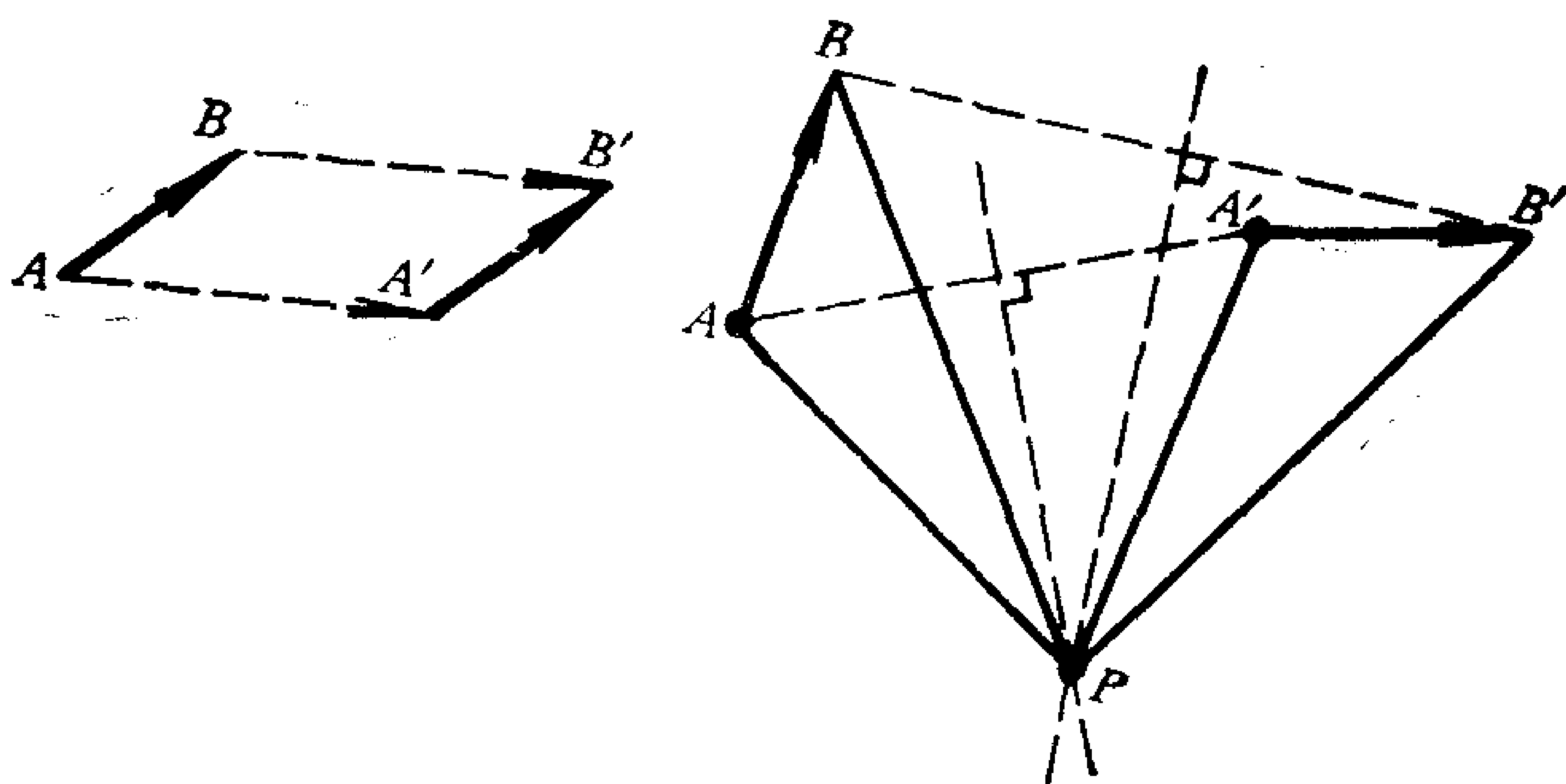


图 18

$\overrightarrow{A'B'}$，且$\overrightarrow{A'B'}$不会与$\overrightarrow{AB}$平行且同向，故有平面上的旋转把$\overrightarrow{AB}$变为$\overrightarrow{A'B'}$。这个旋转就是该合成运动。

更具体地说，设π_a是平移，其中$\boldsymbol{a}$是平移向量；P_φ是以P为中心的旋转，旋转角为φ。显然$P_\varphi\circ\pi_a$仍然是正等距变换。如图19所构造的点Q显然是$P_\varphi\circ\pi_a$的不动点，故$P_\varphi\circ\pi_a$是

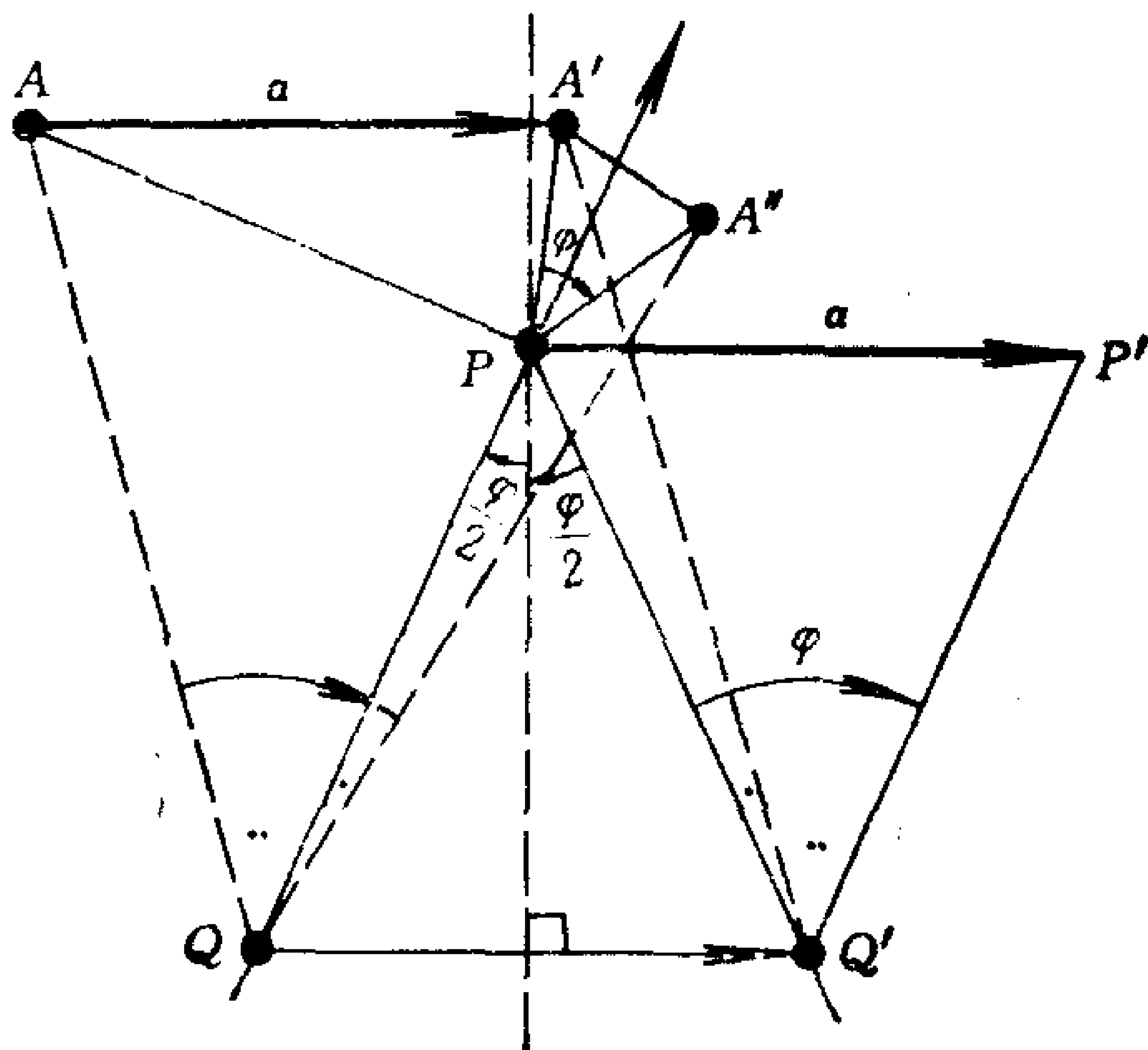


图 19

以Q为中心的旋转．设A是平面上任意一点，并且设$A' = \pi_a(A)$，即$\overrightarrow{AA'} = \boldsymbol{\alpha}$；$P_\varphi(A') = A''$．连结$A'Q'$,$QA''$,$QA$．显然$\triangle PQA'' \cong \triangle PQ'A'$，故$QA'' = Q'A' = QA$，又$\angle PQA'' = \angle PQ'A$,$\angle AQP = \angle A'Q'P'$，故$\angle AQA'' = \angle PQ'P' = \angle Q'PQ = \varphi$，所以

$$A'' = Q_\varphi(A) = P_\varphi \circ \pi_a(A),$$

即$Q_\varphi = P_\varphi \circ \pi_a$．

同理可证：$\pi_a \circ P_\varphi$也是一个角度为φ的旋转．

思考题：$\pi_a \circ P_\varphi$的旋转中心（即不动点）在何处？

(3) 设P,Q是平面上两个不同点，则旋转P_φ与Q_ψ的合成运动或者是一个平移（若$\varphi + \psi = k \cdot 360°$），或者是一个旋转，且旋转角为$\varphi + \psi$（若$\varphi + \psi \neq k \cdot 360°$）．

证明　$P_\varphi \circ Q_\psi$必是平面上的正等距变换．由(1)可知，$P_\varphi \circ Q_\psi$或者是平移，或者是旋转．设$Q_\psi(\overrightarrow{AB}) = \overrightarrow{A'B'}$，$P_\varphi(\overrightarrow{A'B'}) = \overrightarrow{A''B''}$．当$\varphi + \psi = k \cdot 360°$时，$\overrightarrow{AB}$与$\overrightarrow{A''B''}$是平行且同向的有向线段，且$|\overrightarrow{AB}| = |\overrightarrow{A''B''}|$，于是有平移将$\overrightarrow{AB}$变到$\overrightarrow{A''B''}$，该平移就是$P_\varphi \circ Q_\psi$．当$\varphi + \psi \neq k \cdot 360°$时，由图20可求出点$R$，它是$P_\varphi \circ Q_\psi$的不动点，故$P_\varphi \circ Q_\psi$是围绕$R$

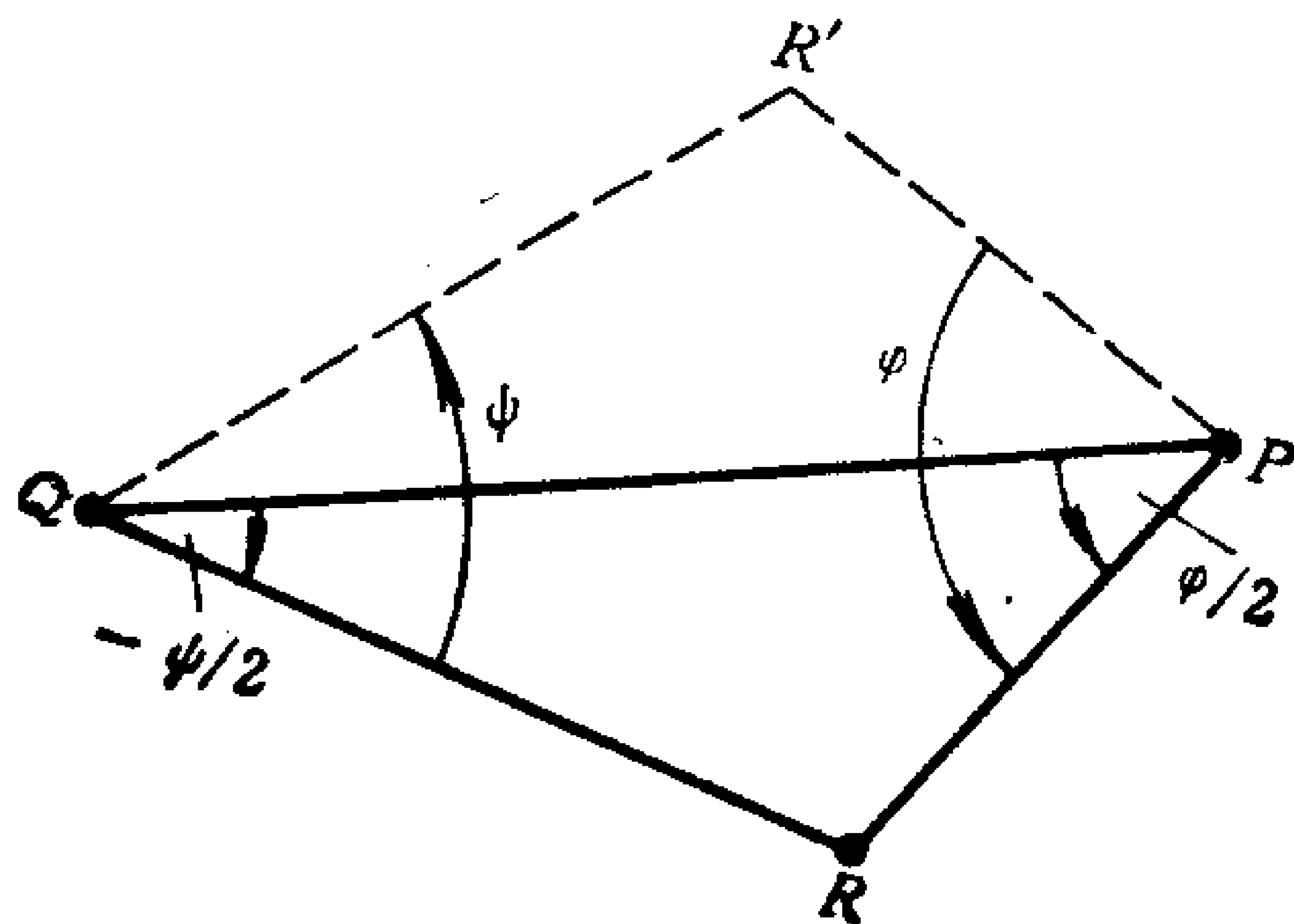


图 20

的旋转。请读者自证：这个合成旋转的旋转角为$\varphi+\psi$。

思考题：旋转$Q_\psi \circ P_\varphi$的中心在何处？

(4) 对于平面上两个正相似图形，必有平面上的平移或旋转相似变换把其中一个图形变为另一个图形。

证明 与事实 (1) 的证明相仿，参看：考克塞特、格雷策著的"几何学的新探索"，p.112，定理 4.8.2 (北京大学出版社，1986年出版)。

(5) 平面上任意两个旋转相似变换的合成或者是一个平移，或者是一个旋转相似变换。

证明 这是事实 (4) 的推论。

(6) 平面上两个中心相似变换的合成或者是一个平移 (此时，两个相似变换的比例系数之积为 1)，或者是中心相似变换。

证明 这是事实 (5) 的特例。我们在这里给出一个简单的几何证明。不妨设 P,Q是平面上两个不同点，P^r,Q^s是分别以P,Q 为中心的相似变换，比例系数分别为r,s。

设$\overrightarrow{AB}$是平面上任意一个有向线段，命$P^r(\overrightarrow{AB})=\overrightarrow{A'B'}$，$Q^s(\overrightarrow{A'B'})=\overrightarrow{A''B''}$。于是$\overrightarrow{AB},\overrightarrow{A'B'},\overrightarrow{A''B''}$彼此平行，且作为自由向量有等式：$\overrightarrow{A''B''}=s\cdot\overrightarrow{A'B'}=rs\overrightarrow{AB}$。故当 $rs=1$ 时$\overrightarrow{A''B''}$与$\overrightarrow{AB}$平行、同向、且长度相等，故有平移把 $\overrightarrow{AB}$ 变到$\overrightarrow{A''B''}$，这个平移就是$Q^s\circ P^r$。若 $rs\neq 1$，则连线 AA'',BB'' 相交于一点R，根据笛沙格定理(用于$\triangle AA'A''$和$\triangle BB'B''$)点R落在PQ的连线上 (图21)。很明显，$R^{rs}(\overrightarrow{AB})=Q^s\circ P^r(\overrightarrow{AB})=\overrightarrow{A''B''}$. 最后我们通过计算得到

$$RQ=\frac{s(r-1)}{rs-1}PQ,\quad RP=\frac{1-s}{rs-1}PQ,\quad \frac{RP}{RQ}=\frac{1-s}{s(r-1)},$$

所以 R 是直线 PQ 上确定的点，与 $\overrightarrow{AB}$ 的取法无关。因此 $R^{rs} = Q^s \circ P^r$。证毕。

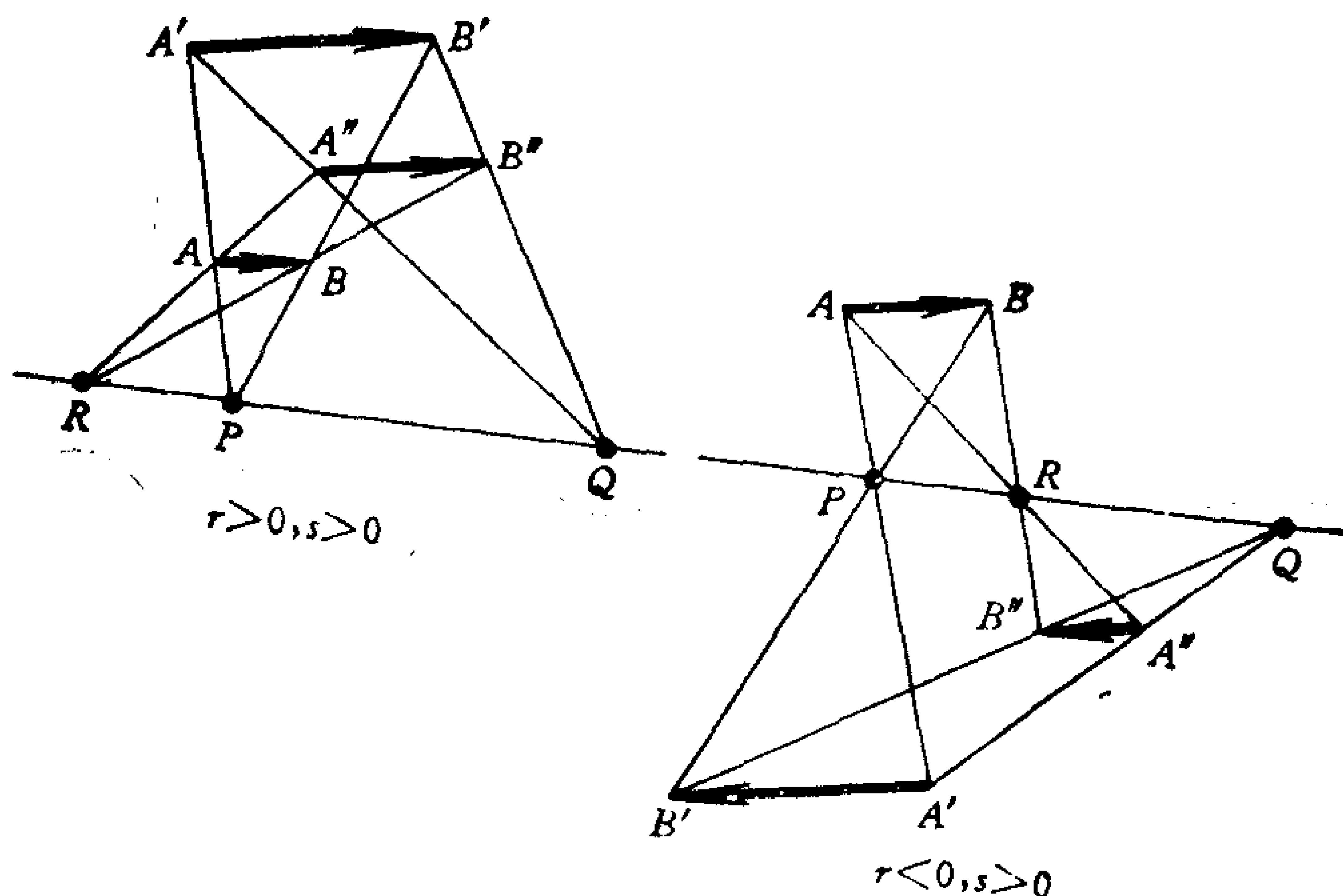


图 21

思考题：能否直接证明 R 是 $Q^s \circ P^r$ 的不动点？

最后我们来观察前文定理 4 的证明中出现的变换 $C_{-a}C^rA^{1/r}A_a$。首先由事实(6)可知 $C^rA^{1/r}$ 是一个平移；由事实(2)知道 $(C^rA^{1/r})A_a$ 是旋转角为 a 的旋转。故由事实(3)推知 $C_{-a}C^rA^{1/r}A_a = C_{-a} \circ [(C^rA^{1/r}) \circ A_a]$ 是一个平移。这里用到了平面上的变换适合结合律。一般说来，变换不满足交换律，例如 $P_\varphi \circ Q_\psi \neq Q_\psi \circ P_\varphi$，$P^r \circ Q^s \neq Q^s \circ P^r$。

顺便提一下：用解析几何方法来证明事实（2）和事实（3）是十分有意思的练习。这时能把旋转中心算出来。这些练习留给读者来完成。

另外，读者还可以考虑：平移与中心相似变换的合成是什么？其不动点是否存在？若有，则如何确定？

几何极值问题①

G. D. CHAKERIAN, L. H. LANGE

1. 引言

微积分课程中的一个典型的练习是：给定一个高为 a，底边为 b 的三角形，要找内接于这个三角形、其一边沿着三角形的底且具有最大面积的矩形。

对于一个学生来说，如果他偶然看一下这个问题的避开微积分的另一种解法，那么至少会开拓他的眼界。上述问题可以用一个初等的不等式解决如下：

如图 1 所示，给定的三角形有两种本质上不同的可能形式。在情形（a）中，顶点 C 是底边“上方”的某个点；在

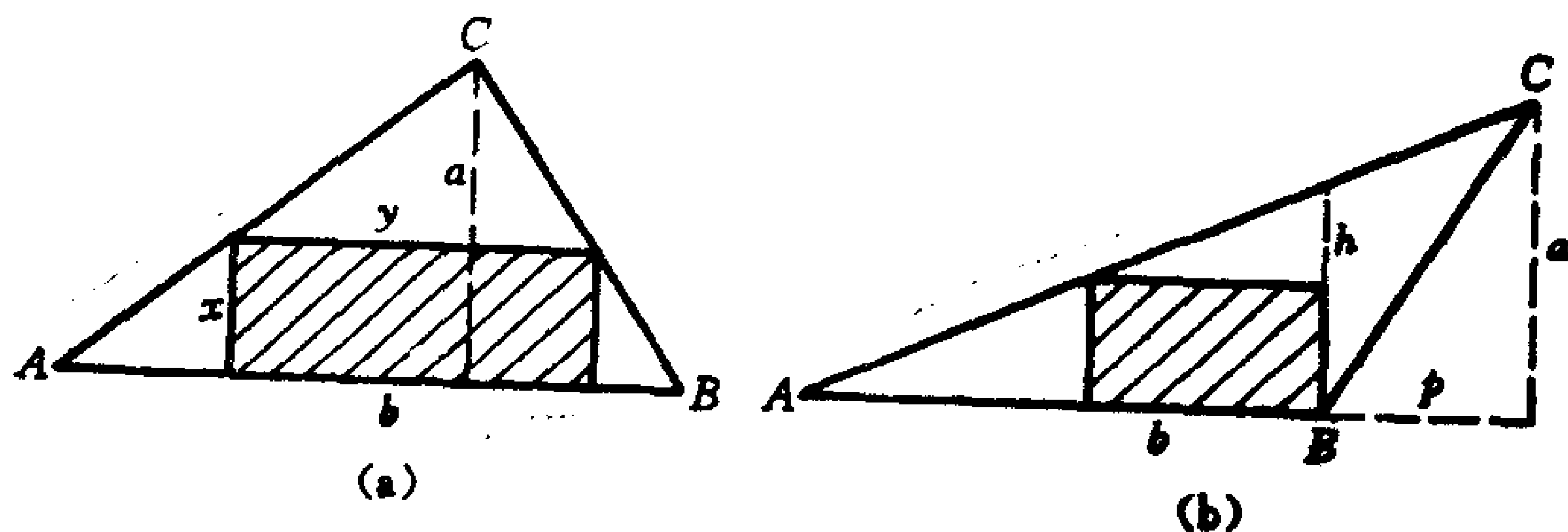


图 1

① GEOMETRIC EXTREMUM PROBLEMS, *Math. Magazine*, Mar.-Apr. (1971), 57—69.

情形（b）中，顶点C不是位于那样的地方。我们先解情形（a），而情形（b）的解很容易由情形（a）的解来确定。

用图1中的符号，我们寻找乘积xy的最大值，由相似三角形，有$y=(b/a)(a-x)$。于是，$xy=(b/a)(x)(a-x)$，我们需要寻找x的值，$0\leqslant x\leqslant a$，使这个量达到最大。为寻找$x(a-x)$的最大值，与其用导数，不如由直接观察得出，注意到

$$x(a-x)=(a/2)^2-\{x-(a/2)\}^2,$$

当且仅当$(x-a/2)^2=0$，也就是$x=a/2$时达到最大。因此，

$$xy=(b/a)(x)(a-x)\leqslant\frac{1}{2}(ab/2)=\frac{1}{2}\text{面积}(\triangle ABC),$$

当$x=a/2$时等式成立。于是，最大的矩形的高为$x=a/2$而面积恰为给定三角形面积的一半。在情形（b）中，最大的矩形的高为$h/2$而面积小于给定三角形面积的一半（在这种情形中，最大面积是$\frac{1}{2}\{b/(b+p)\}\cdot$面积$(\triangle ABC)$）。

注意，在情形（b）中，如果我们选取AC为底边，在其上放置矩形，我们将得到一个其面积是给定三角形面积一半的最大矩形。因此，我们看到在任何情况，都能够使面积为其一半的某个矩形内接于给定的三角形。

在微积分教科书中，或明或暗地通常只处理那些假设成某种特殊位置的最佳图形。例如，在上述的问题中仅考虑其边位于给定三角形底边上的矩形。但是很自然要问：包含在给定三角形内的所有矩形中，哪一个具有最大面积？这个最大面积是否会大于给定三角形面积的一半。

我们将在第2节中回答这个问题（见下面的定理3）而

在较后的一些节中将考察另一些极值问题，在那里最佳图形通常都假设成在某种特定的特殊位置下。然而，在这篇论文中我们的主要目的是对于上述类型的一般几何极值理论提供一种容易理解的描述，以及一些在教学中有用的例子（不一定只在微积分教程中）。在下面，基本的课题是用仿射变换来简化这种类型的问题。

2. 外接于凸集有最小面积的多边形

我们将涉及平面凸集，即具有下面性质的平面点集：连接点集中任何两点的线段仍包含在点集中。而凸区域就意味着是紧的（即，闭的和有界的）且有非空内部的平面凸集。

众所周知，如果K是凸区域，$n\geqslant3$是给定的整数，那么至少存在一个包含K有最小面积的凸n边形。显然，这样一个n边形必须外接于K——即它的边必须与K的边界相交。下面的定理描绘了使我们感兴趣的特性。

定理1 设K是一个凸区域，$n\geqslant3$是给定的整数。设P是包含K的具有最小面积的凸n边形。那么P的边的中点必在K的边界上。

注释 虽然这是一个有名的结论（见[3]p.6），我们知道要达到发表证明的程度并不容易，因此我们感到在第4节中给出一个初等的证明也是正当的。这个证明将说明在第3节中所阐述的仿射变换方法的有效性。

作为定理1的一个简单应用，考虑K是平行四边形的情形。而设T是包含K具有最小面积的三角形。按照定理1，T的边的中点碰到K。这仅当K的一条或两条边位于T的边上，其相互的位置如图2(a)或(b)所示的那样时才可能，

读者很容易承认这一结论。

注意，在任何情形，面积$(T)=2\cdot$面积(K)。记住T是最小的三角形，于是有下列结果：

定理2 设K是一个给定的平行四边形，T是任何一个包含K的三角形。那么，面积$(T)\geqslant 2\cdot$面积(K)，而等式成立当且仅当T是如图2所示的那种特殊位置的三角形。

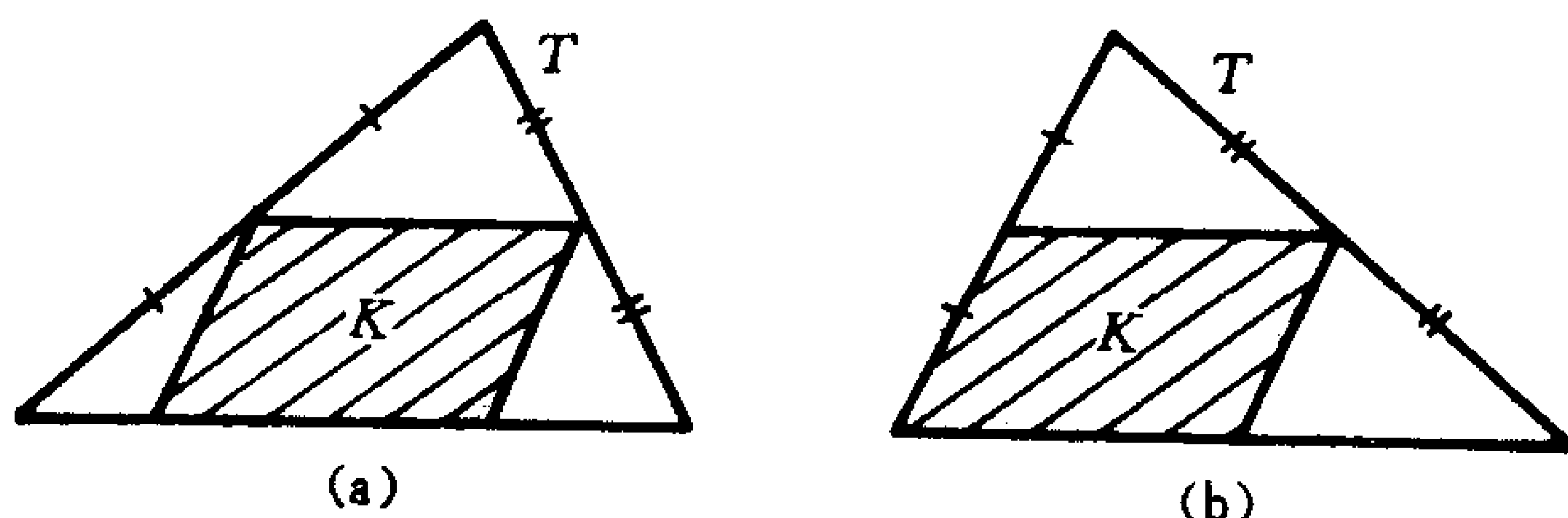


图 2

现在我们能够解决在第1节中所提出的问题了。我们有

定理3 设T是一个给定的三角形，而R_0是包含在T中具有最大面积的矩形。那么，面积$(R_0)=\frac{1}{2}$面积(T)，而R_0是一种特殊位置的矩形，即它的一条边位于T的一条边上，且T的另两个边的中点是R_0的顶点。

证明 假设R是包含在T中的一个矩形但不是在上述规定的那种特殊位置的矩形。那么，定理2意味着T不是包含R具有最小面积的三角形；因此，如果T^*是这样一个最小三角形，就有

$$\text{面积}(R)=\frac{1}{2}\text{面积}(T^*)<\frac{1}{2}\text{面积}(T)。$$

另一方面，如果R_0是在上述规定的那种特殊位置下的矩

形，那么面积$(R_0)=\frac{1}{2}$面积(T)。从而R_0是包含在T中具有最大面积的矩形（注意，如果T是锐角三角形，这样的矩形有三个，如果T是直角三角形那么就有两个，如果T有一个钝角那就只有一个）。证完。

注释 导致定理3的问题是由[7]的作者之一在一篇论文中提出的。同时，M.T.Bird([1])给出了定理3的一个简捷的证明。

定理2包含在Fulton和Stein的一个结果中[4，定理1]。

平行四边形在它们关于外接三角形的状态上是稍微有些极端——不能够保持在其面积小于它的面积两倍的三角形内部。很自然会提出两个问题：

(1) 平行四边形是否是仅有的以这种（可悲的）方式处身的凸区域？

(2) 是否每一个凸区域K都包含在其面积小于或等于两倍K的面积的某个三角形内？

在阐述了某些工具，也就是能帮助我们简化这些和与其有关的问题的仿射变换之后，我们将转向这些问题。

关于定理3，C.Radziszewski([9])证明了每个凸区域K包含一个面积是它一半的矩形。

3. 仿射变换

在这一节中，我们要注意平面非奇异仿射变换的某些有用的性质。所谓非奇异仿射变换，它是(x,y)-平面到自身的变换，把每个点(x,y)变到点(x',y')使得

$$x'=ax+by+e,\qquad y'=cx+dy+f,$$

其中$a,b,\cdots,f$是给定的实数且满足条件$ad-bc\neq 0$。由于我们仅关心非奇异变换，将始终用术语"仿射变换"意指非奇异仿射变换。

在后面将发现仿射变换是很有用的，在列出它的性质之前，我们先看一个很有启发性的例子，这个例子涉及这种类型的函数并考虑某种几何极值问题（对于与之有关的讨论，见[8]）。

设a和b是给定的实数，满足$0<a\leqslant b$，又设$\mathscr{P}$是全体实数对(x,y)所组成的集合——平面，考虑函数μ，它映射$\mathscr{P}$到$\mathscr{P}$，且把点(x,y)变到点$(x',y')=(ax,by)$。如果我们现在考虑全体使得$x^2+y^2\leqslant 1$的点(x,y)的集合D，可以看到对应点(x',y')必须满足$(x'/a)^2+(y'/b)^2\leqslant 1$。

函数μ是一个很特殊的仿射变换；事实上，它是线性变换的一个简单例子。等式$x'=ax$和$y'=by$告诉我们，在$\mathscr{P}$中任意给定(x',y')，在$\mathscr{P}$中存在唯一的(x,y)使得μ把(x,y)变到(x',y')；也就是说μ是一对一的和满的函数。特别地，我们看到如图3所示的（闭的）椭圆盘的每一点都是

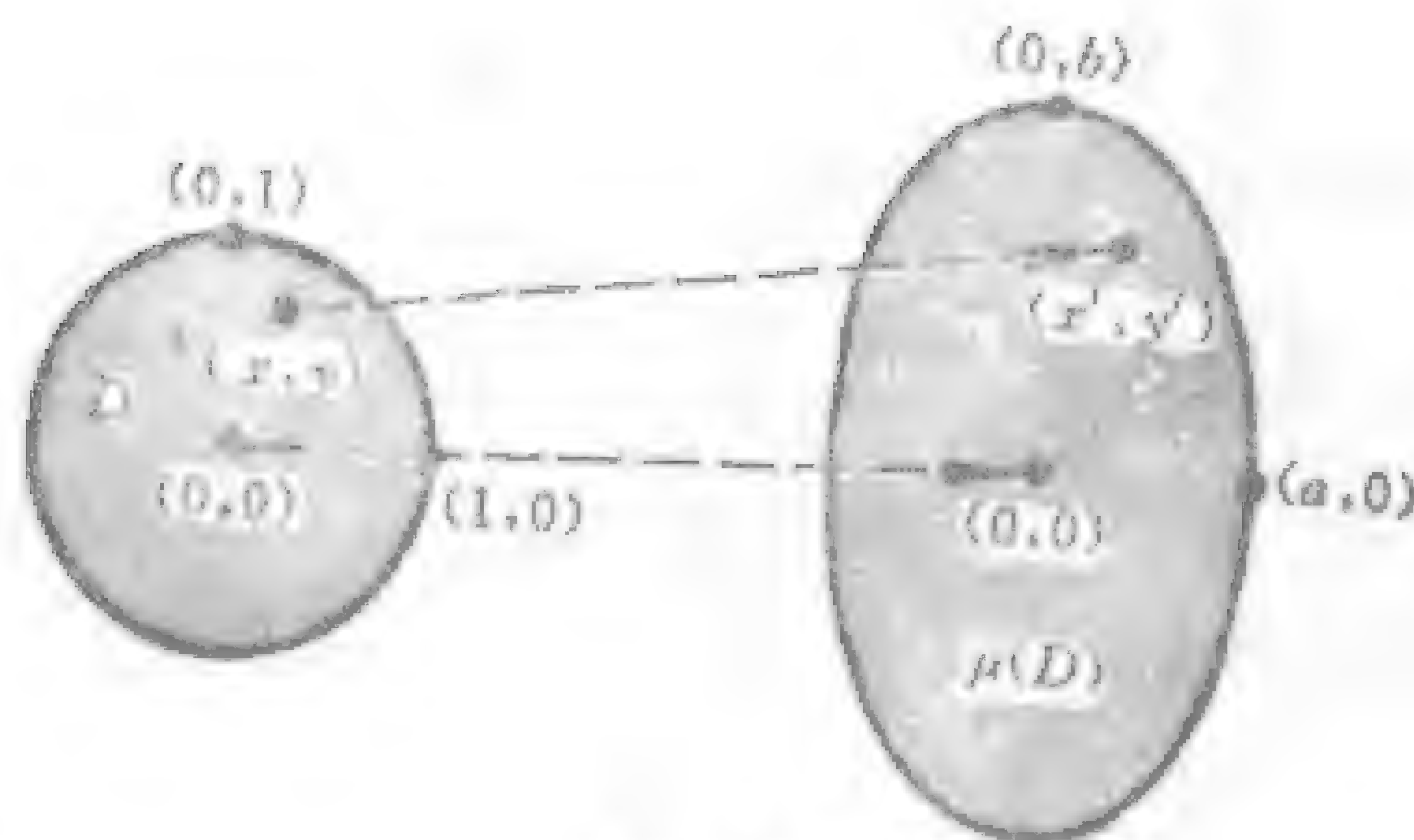


图 3

(闭的) 圆盘D中的恰好一个点的像。函数μ“拉开圆片D，没有任何折叠地(由于μ是一对一的)完全(由于μ是满的)而恰好覆盖椭圆盘$\mu(D)$。”

现在假设p,q和r是给定的实数，考虑集合

$$\{(x,y):px+qy+r=0\},$$

这是一条直线。然后函数μ把这个集合变为集合

$$\left\{(x',y'):\frac{p}{a}x'+\frac{q}{b}y'+r=0\right\}.$$

于是，在映射μ之下，任何给定的直线的像也是直线。容易看到，连结(x_1,y_1)和(x_1,y_2)的垂直线段的像是连结像点(ax_1,by_1)和(ax_1,by_2)的垂直线段(见图4)。我们还要注意，像线段的长度是$b\cdot|y_2-y_1|$，当然这里$|y_2-y_1|$是原线段的长度。类似地，水平像线段的长度是原水平线段长度的a倍。

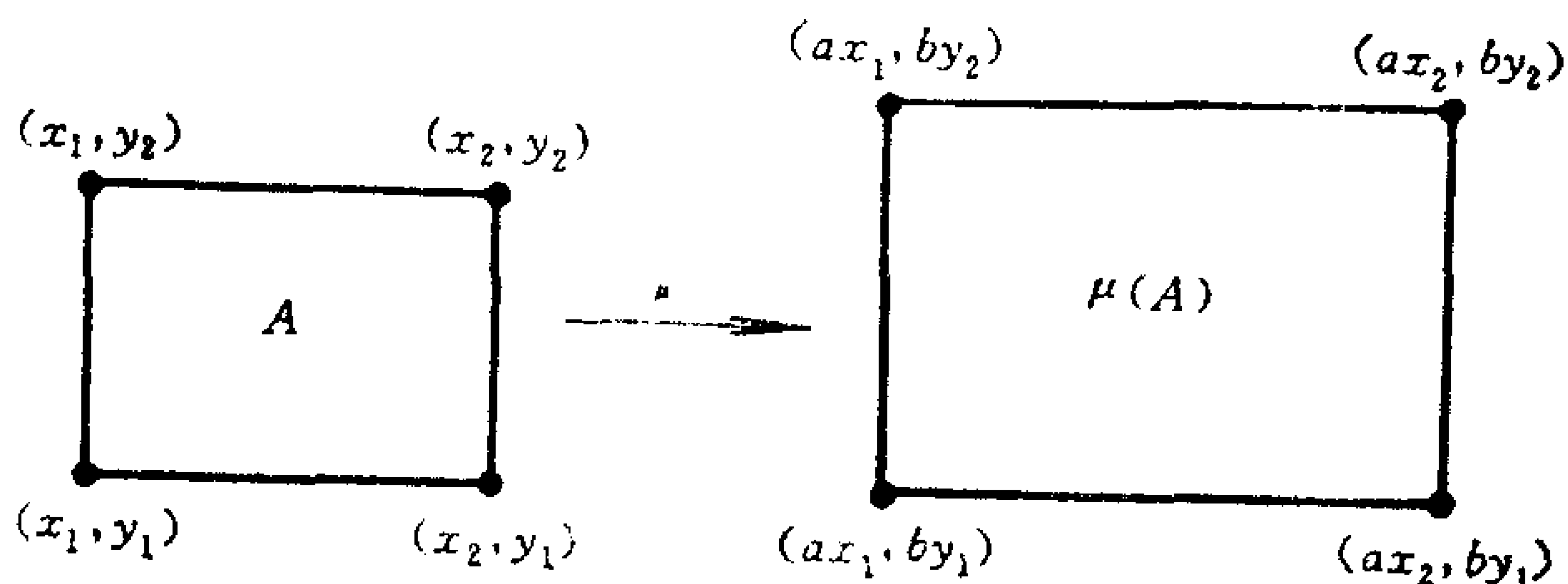


图 4

现在，我们再看在函数μ映射下圆盘D的像，将看到如果图4中的矩形A在圆盘D内，那么它的像$\mu(A)$是一个矩形且在椭圆盘$\mu(D)$内。所以，如果A的面积是α，那么矩形

$u(A)$的面积是 $(ab)\times a$。

由此可见，如果$\sum a_i$是D内这样的矩形的有限集合的总面积，那么 $(ab)\sum a_i$就是在椭圆盘 $\mu(D)$内相应的矩形集合的面积。

因此，如果把任意有限个这样的矩形（各种尺寸的）“填塞”到D中，当然要求这些矩形除沿边界外不互相“覆盖”，我们知道它们的面积的和$\sum a_i$满足$\sum a_i<4$，这里 4 是包围D的边长为 2 的正方形面积。因此，在$\mu(D)$内相应的矩形面积之和满足 $(ab)\sum a_i<(ab)(4)$。

对于全体可能的数$\sum a_i$的集合来说，数 4 并不是最小的数而是一个上界。当然这个荣誉是被数π所享有的；因为π是所有这样的可能和数$\sum a_i$集合的最小上界，它被称为圆盘D的面积。

由此可见，由于$(ab)\pi$是所有数$(ab)\sum a_i$的集合的最小上界，$(ab)\pi$就一定是椭圆盘$(x/a)^2+(y/b)^2\leqslant 1$ 的面积。

练习1 证明椭球$(x/a)^2+(y/b)^2+(z/c)^2\leqslant 1$ 的体积是 $\frac{4}{3}\pi abc$。

我们现在列出（不加证明）将要用到的平面仿射变换的性质。

性质1 如果l和m是平行线，那么它们的像也是平行线（“平行线变到平行线”）。

性质2 如果u^*和v^*分别是位于两平行线上的线段u和v的像，那么

$$\{长度(u^*)/长度(v^*)\}=\{长度(u)/长度(v)\}$$

（“保持平行线段的长度比”）。

特别地，我们有

性质3 如果u^*是线段u的像，那么u^*的中点是u的中点的像（“中点变到中点”）。

性质4 任何凸区域的像也是凸区域。

性质5 如果U^*和V^*分别是区域U和V的像，那么

$$\{面积(U^*)/面积(V^*)\}=\{面积(U)/面积(V)\}.$$

（“保持面积的比”）

性质6 任何给定的三角形都可用某个适当的仿射变换映射成任何另一个给定的三角形。（注意性质3就意味着它们的质心也是对应的）

性质7 任何平行四边形都可用某个仿射变换映射成任何另一个给定的平行四边形。

性质8 任何椭圆的像也是一个椭圆，而且任何一个椭圆都可用某个仿射变换映射成任何另一个椭圆。（注意，椭圆的中心变到其像的中心。确实，性质3意味着中心对称图形总是变到中心对称图形，且中心是相应的）

性质9 任何（非奇异）仿射变换是连续的和可逆的，且其逆也是（非奇异）仿射变换。

这里是这些性质的一个简单推论：

引理 在任何三角形T中都有一个且仅有一个椭圆E_0在T的各边中点与T的边相切。

证明 至少有一个这样的E_0存在.这是由于我们可以仿射地把T变换到一个等边三角形T^*并考虑T^*的内切圆（称之为E_0^*）．在逆变换下E_0^*的像就是所需要的椭圆E_0。

假设F_0是在T内且在T的各边中点与T的边相切的另一个椭圆。那么，F_0在上述变换下的像F_0^*是与等边三角形T^*

的边在它们的中点相切的椭圆。但F_0^*的中心与T^*的质心重合（为了看清这一点，将F_0^*映射到一个圆并注意到T^*一定映射到一个外切于此圆的等边三角形——然后用性质6和性质8）．因此，通过T^*的质心的中心反射把F_0^*变到它自己．从而F_0^*不仅包含T^*的边的中心，而且也包含T^*通过其质心的反射的边的中点。由于这些中点是内接于E_0^*的正六边形的顶点，因为一个椭圆可被五个点所决定，从而$F_0^* = E_0^*$；因此$F_0 = E_0$。证完。

这个引理使我们能够建立一个类似于定理3的极值性质。

定理4 任何给定的三角形T都包含有唯一的一个具有最大面积的椭圆E_0.这个椭圆与T的边在它们的中点相切，且面积$(E_0) = (\pi/3\sqrt{3})\cdot$面积$(T)$。

证明 假设E是包含在T中的一个椭圆，但T的边的中点并不碰到E，设E_0是唯一的在T内部与T的边在它们的中点相切的椭圆。进一步设S是包含E且具有最小面积的三角形．那么，根据定理1，S的边的中点必在E上，而面积$(S)<$面积(T)。

现在仿射地映射S到T。那么E也就映射到在T内与T的边在它们的中点相切的椭圆E^*；因此，根据上述引理，$E^* = E_0$。根据性质5，就有

$$\frac{\text{面积}(E)}{\text{面积}(S)} = \frac{\text{面积}(E^*)}{\text{面积}(T)} < \frac{\text{面积}(E_0)}{\text{面积}(S)},$$

因此，面积$(E)<$面积(E_0)。于是，E_0是包含在T内具有最大面积的唯一的椭圆。把T映射到等边三角形就得到比值

$$\{\text{面积}(E_0)/\text{面积}(T)\}=\pi\sqrt{3}/9.$$

证完。

注释 定理4是下面一般结果的一个特例：每一个凸区域都包含唯一的具有最大面积的椭圆，同时也包含在唯一的具有最小面积的椭圆内。对于这一结果的讨论并推广到高维空间，读者可以参考[2]。

练习2 用这一节的方法证明任何给定的三角形T都包含在唯一的具有最小面积的椭圆E_1中，而面积$(E_1)=4\pi/3\sqrt{3}\cdot$面积$(T)$。[提示：用仿射变换把$T$映射到一个等边三角形$T^*$并设$E_1^*$是$T^*$的外接圆。所要的$E_1$是$E_1^*$在逆变换下的像。]

练习3 用练习2和定理4，证明任何三角形的外接圆半径至少两倍于内切圆半径（对于这个性质的一个很漂亮的证明是由I. Ádám给出的，见[3,p.28]。）

4. 定理1的证明

为证明定理1，需要下述引理，这个引理将告诉我们如何通过一个角内的给定点用一条直线截出具有最小面积的三角形。

引理 设XOY是一个给定的角(见图5)。那么，

(1) 对于角内部的每一点M，存在且仅存在一条过M在$\overrightarrow{OX}$上的端点为A，在$\overrightarrow{OY}$上的端点为B且被M平分的线段$\overline{AB}$。而且

(2) 从角中被通过M的线所截出的所有三角形中，$\triangle AOB$是唯一具有最小面积的。

(3) 设$\triangle(Q)$表示与角XOY内部的点Q联系着的最小三

角形．如果沿$\overline{AB}$，$Q\to M$，那么$\triangle(Q)\to\triangle(M)$．依此意味着$\triangle(Q)$的顶点趋向于$\triangle(M)$的相应的顶点．

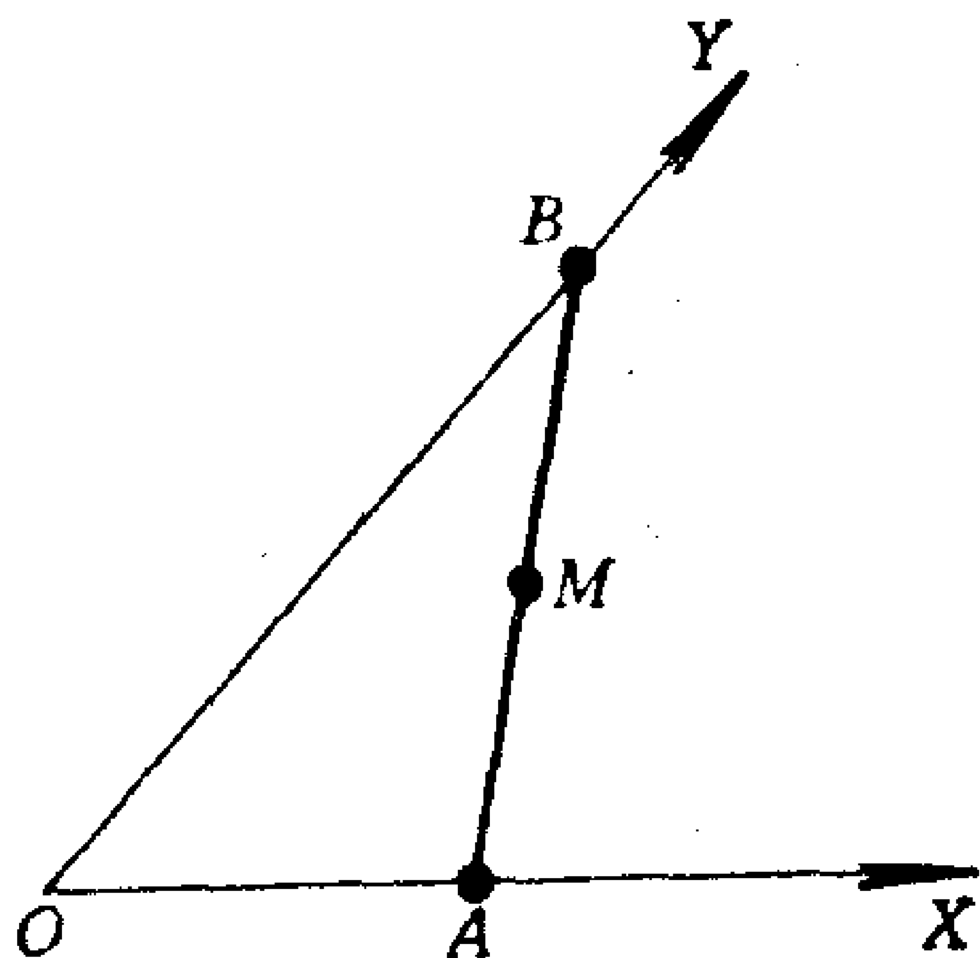


图 5

证明 一个简单的连续性论证就证明了总存在某个被M平分的线段$\overline{AB}$．现在用一个仿射变换把A变到$A^*=(1,0)$，O变到$O^*=(0,0)$，B变到$B^*=(0,1)$，并考虑这个仿射变换对$\triangle AOB$变换的形状．那么自动地M变到$M^*=(1/2,1/2)$．现在我们可以在这种位置下完全证明这个性质．为此恢复原图中的符号而省略星号．

事实上，设$Q=(\xi,1-\xi)$，$0<\xi<1$，是AB内部的任一点，而$\triangle COD$是由通过Q的直线截出的任何三角形．然后，设$C=(c,0)$，有

$$\text{面积}(\triangle COD)=\frac{c^2(1-\xi)}{2(c-\xi)}\geqslant 2\xi(1-\xi)$$

（不等式是$(2\xi-c)^2\geqslant 0$的结果）．等式当且仅当$c=2\xi$时成立．因此，有唯一的面积减到最小的三角形$\triangle(Q)$，它有一边通过Q且被Q平分，而且

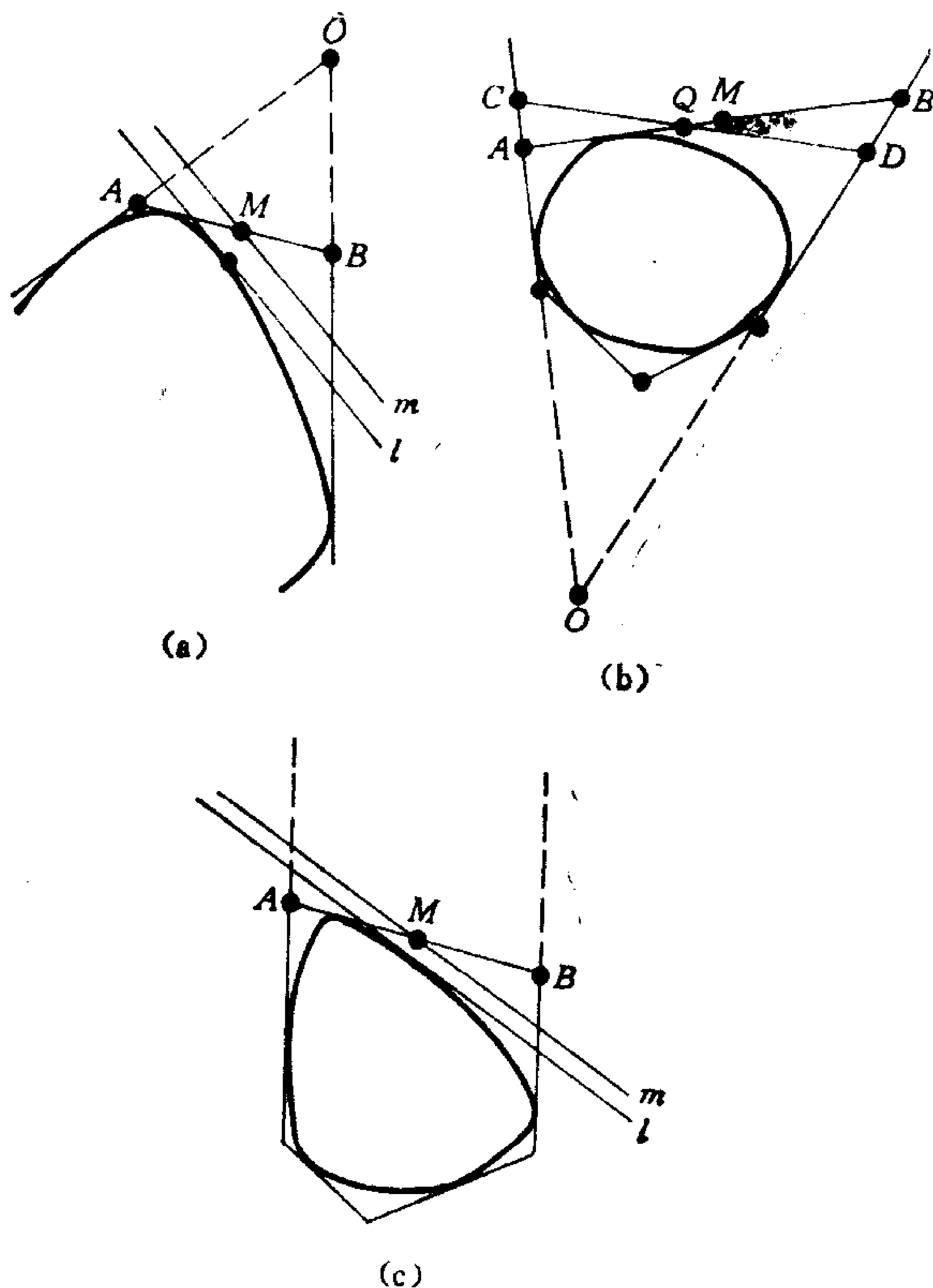


图 6

$$\text{面积}(\triangle(Q)) = 2\xi(1-\xi)。$$

如果$A'OB' = \triangle(Q)$，那么注意 $A' = (2\xi, 0)$ 和 $B' = (0, 2-2\xi)$．于是，当沿$\overline{AB}$，$Q \to M$时，就有 $A' \to A = (1,0)$ 和$B' \to B(0,1)$；那就是$\triangle(Q) \to \triangle(M)$．将第3节中仿射变换的性质3，性质5和性质9应用于上面曾用过的变换的逆上

就很容易证明了引理中所断言的性质(1),(2)和(3)。

现在可以来证明定理1了.假设P是外切于凸区域K的有最小面积的n边形，并假设某一边$\overline{AB}$的中点M不碰到K。我们考虑图6中所示的三种可能的情形。

在情形(a)中，P的相邻于$\overline{AB}$的边，当延长时相交于点O，使得$\triangle AOB$不包含K。在情形(b)中，$\triangle AOB$包含K.在情形(c)中，相邻于$\overline{AB}$的两边是平行的。我们将证明在每一种情形中都可能构造包含K而面积比P小的多边形;那么，这个矛盾就证明了最小n边形每一边的中点确实一定在K上。

在情形(a)中,由于M在K的外面，总存在把M和K隔离开的K的一条支撑线l(K的支撑线L是使K位于L的一条边上且$K\cap L\neq\varnothing$的直线)。设m是过M平行于l的直线(图6(a)).注意到，根据刚才的引理，m截出一个其面积严格大于面积($\triangle AOB$)的三角形。而l截出一个面积甚至更大的三角形;因此，用l和P就可以引进一个包含K且面积比P小的n边形。

在情形(b)中，可以利用引理的性质(3)引进一条线段$\overline{CD}$，它的中点Q在$\overline{AB}$上，$Q\neq M$，且$\overline{CD}$不与K相交(如图6(b)所示)。再根据引理，$\triangle COD$的面积严格地小于所有其它的由通过Q的直线截出的三角形的面积;因此，面积($\triangle COD$)<面积($\triangle AOB$).那么显然我们可以引进一个包含K而其面积比P小的n边形。

在情形(c)，先取(两平行边的)中线与K的边界的交点(靠近M)，l就取成在这一点上K的一条支撑线。然后m是过M与l平行的直线。容易看出，用m和P可得到一个包含K其面积与P相等的n边形，用l就得到一个有更小面积的n边形。证完。

5. 包含一个凸区域的最小三角形

我们现在来回答在第 2 节最后提出的一个问题。

定理5 每一个凸区域K都包含在至多两倍于它的面积的某个三角形中。

证明 如果T_0是包含K而有最小面积的三角形，且面积$(T_0)\leq 2\cdot$面积(K)，那就足够证明这一结论了．为了做到这一点，设T_0是外切于K的最小三角形，在K上的它的边的中点为A,B,C。如图7所示，用画K的平行于T_0的边的支撑线而得到的T是相似于T_0的三角形，并设A',B',C'是T的边与K的交点。

我们现在证明六边形$AB'CA'BC'$有至少两倍于$\triangle ABC$的面积。那么就有

$$\text{面积}(K)\geq\text{面积}(AB'CA'BC')$$

$$\geq 2\cdot\text{面积}(\triangle ABC)=\frac{1}{2}\text{面积}(T_0),$$

而这就证明了定理．

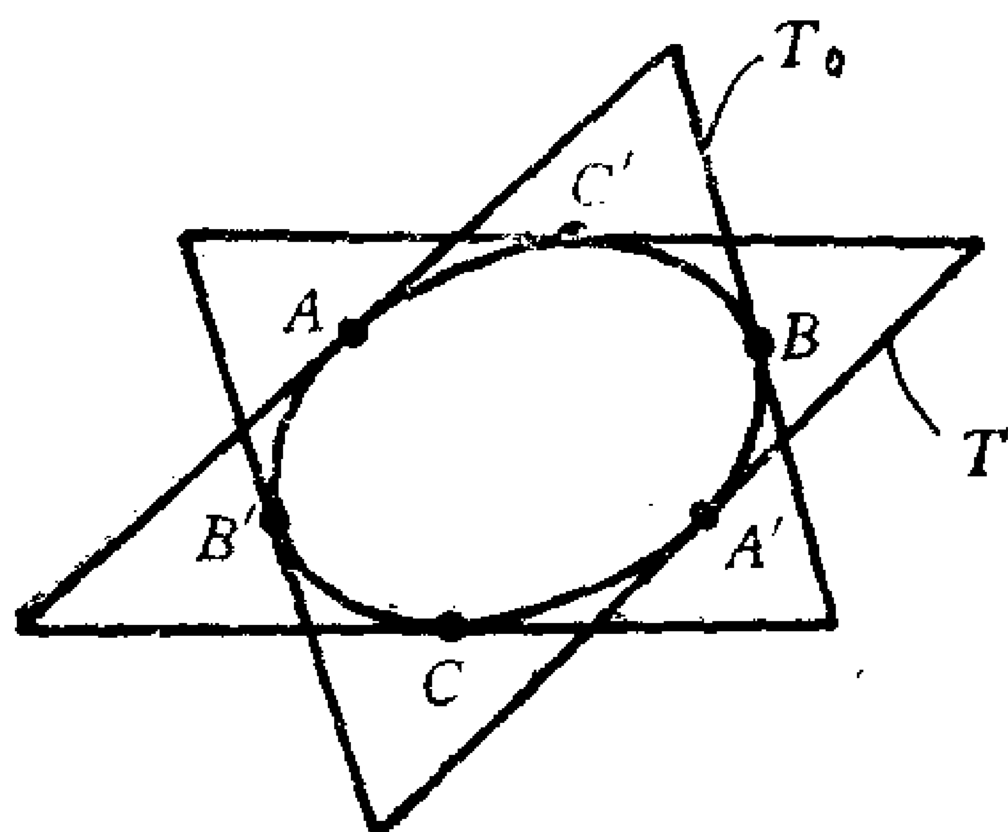


图 7

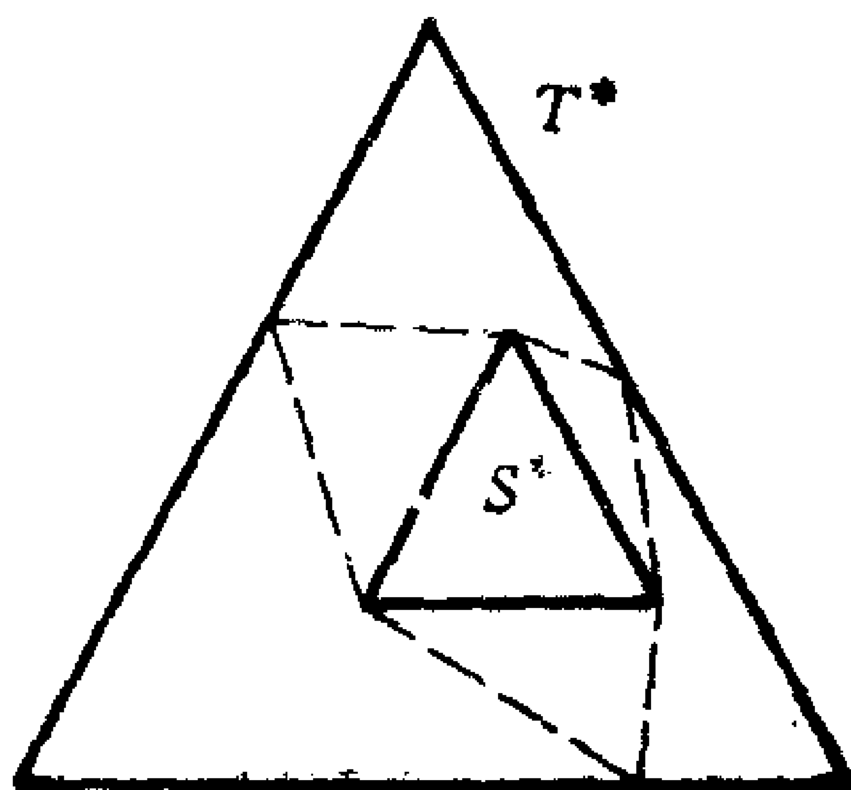


图 8

现在我们仿射地变换图T中的外形，使T被映射成一个等边三角形T^*。然后，$\triangle ABC$被映射到T^*内部的等边三角形S^*，它的边平行于T^*的边，而六边形$AB'CA'BC'$被映射成如图8中虚线所示的六边形。

由于T_0是最小的并与T相似，T的每一边至少像T_0的相应边一样长，因此也就至少两倍于$\triangle ABC$的平行边的长度。于是，T^*的边至少是S^*边长的两倍。由于仿射变换保持面积比，我们仅需要证明在这些条件下图8中虚线六边形至少有S^*的面积的两倍。为了证明这一点，从S^*的质心O引T^*的边的垂线，作出图9中的虚线六边形。

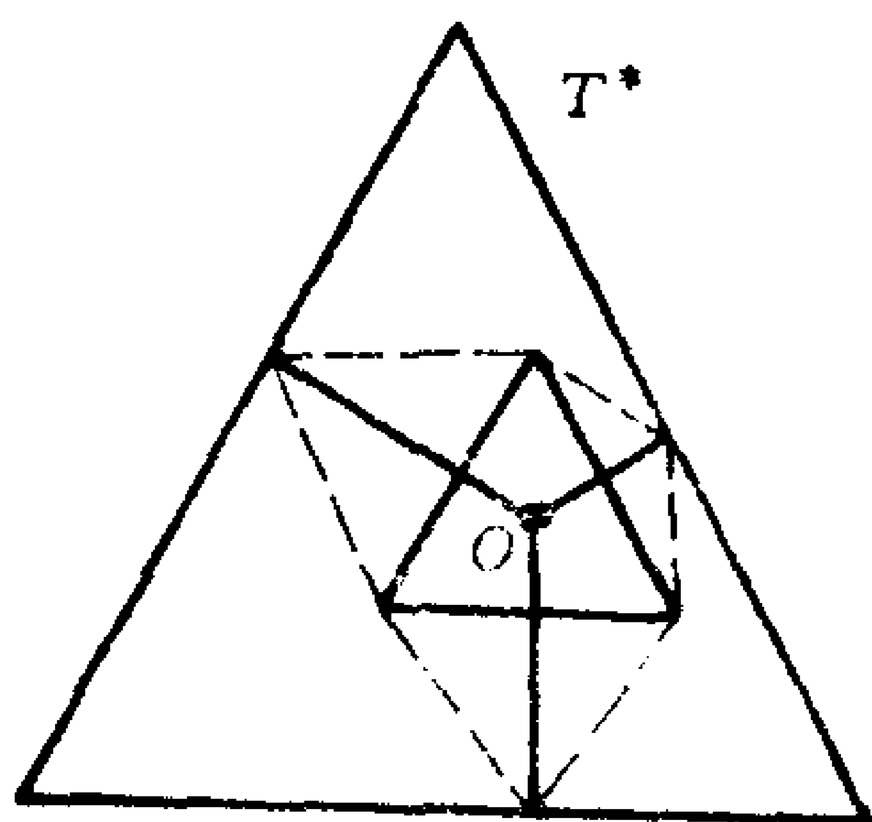


图 9

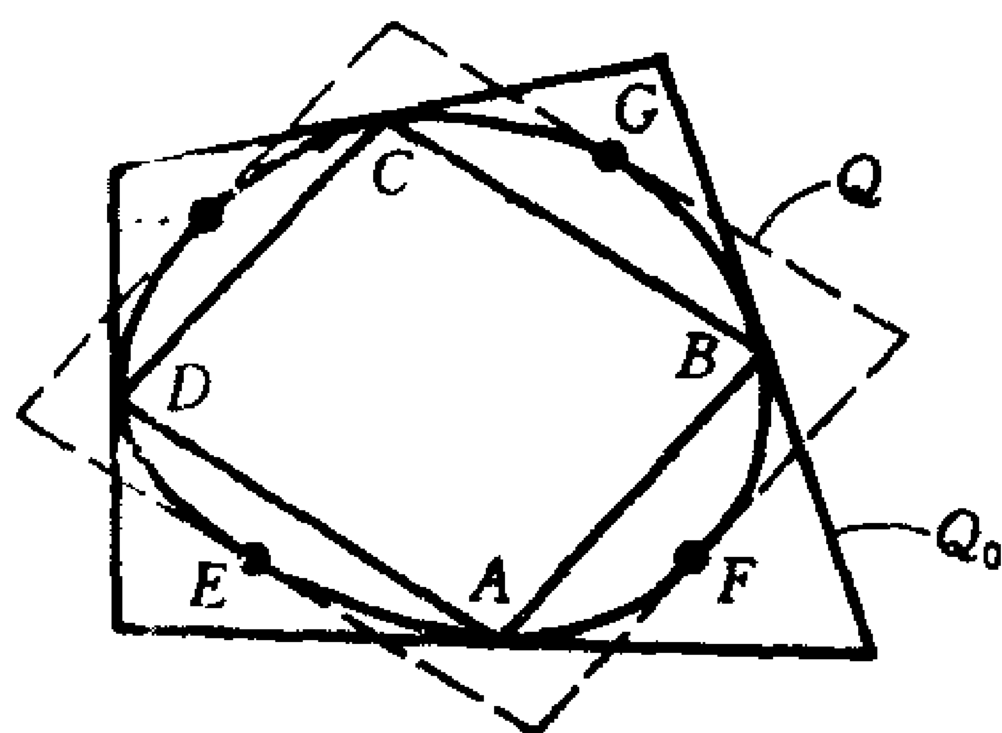


图 10

图9中虚线六边形有着与图8中虚线六边形同样的面积。我们把剩下的作为练习留给读者去证明这个六边形至少有两倍于小三角形S^*的面积[提示：从等边三角形内任一点所作各边垂线的长度和总等于三角形的高。再记住大三角形有着至少两倍于小三角形边长的边]。于是，读者就完成了证明.

注释 定理5是首先由Gross证明的（[5]）。Gross还证明了，如果包含K的最小三角形恰有K的面积的两倍，那么

K一定是平行四边形。这就肯定地回答了在第2节最后提出的问题(1)。

很自然，对于$n>3$的n边形，也可给出定理5的类似结果。下面关于$n=4$情形的部分结果看来是新的：

定理6 每一个凸区域K都包含在一个四边形Q_0中而使得

$$面积(Q_0)\leqslant(\sqrt{2})\cdot面积(K)。$$

证明 设Q_0是包含K具有最小面积的四边形，在K上其各边的中点分别为A,B,C,D。众所周知，$ABCD$是面积为Q_0一半的平行四边形。在图10中用虚线画出外切于K的平行四边形Q，它的各边与$ABCD$的边平行且在点E,F,G,H与K相交。

如果我们能证明顶点为$AFBGCHDE$的八边形Z满足

$$面积(Z)\geqslant\sqrt{2}面积(ABCD),$$

那么就有

$$面积(K)\geqslant面积(Z)\geqslant\sqrt{2}面积(ABCD)=\frac{\sqrt{2}}{2}面积(Q_0),$$

而定理也就跟着得到了。

我们注意到，如果让E,F,G和H沿着Q的它们各自所在的边移动，Z的面积是不变的，而且考虑Q和$ABCD$是矩形的情形就足够了（用一个适当的仿射变换）。换句话说，在如图11所示的情况证明面积$(Z)\geqslant\sqrt{2}\cdot$面积$(ABCD)$就足够了。

在图11中，E,F,G和H是从$ABCD$的中心作Q各边的垂线的垂足。虚线表示的多边形是新的Z。如果s和t是$ABCD$的边长，而s'和t'是Q的相应的平行边的边长，那么容易验证

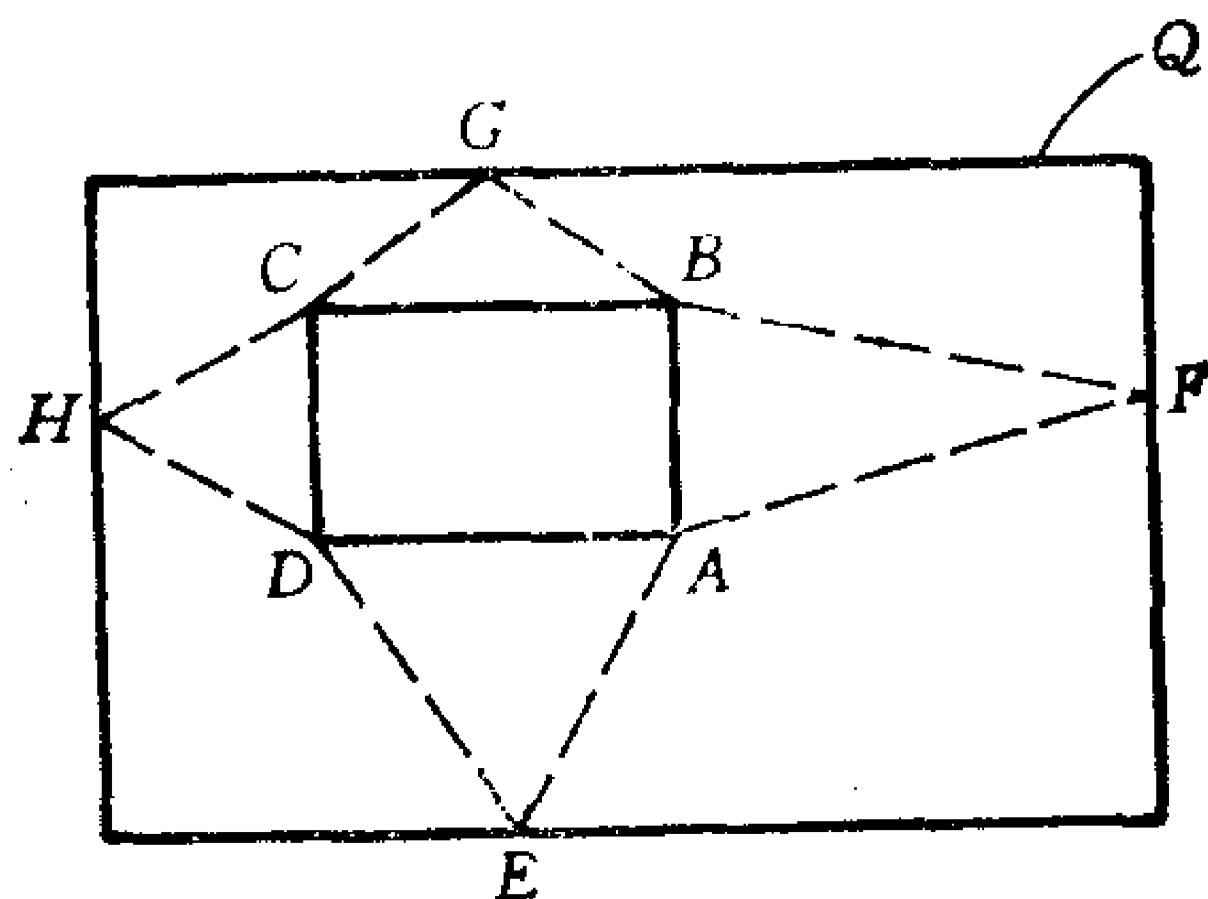


图 11

$$面积(Z)=\frac{1}{2}(st'+s't).$$

回想在最初的形状中有面积$(Q)\geqslant$面积$(Q_0)=2\cdot$面积$(ABCD)$，我们看到$s't'\geqslant 2st$。于是，应用两个数的算术平均值总大于或等于它们的几何平均值这一事实，得到

$$面积(Z)=\frac{1}{2}(st'+s't)\geqslant\sqrt{st's't}$$

$$\geqslant(\sqrt{2})st=\sqrt{2}面积(ABCD),$$

而定理也就得到了。

注释 我们不知道是否定理6已是最好的结果了.有可能每一个K包含在某个四边形Q中使得面积$(Q)\leqslant\lambda\cdot$面积(K)，这里$\lambda<\sqrt{2}$。需要回答的问题为：是否有一个凸区域K，它的所有的外切四边形都有至少是$(\sqrt{2})$面积(K)的面积？

Fejes Tóth ([3,p.38]) 指出对于$n>3$的n边形相应问题的答案是未知的。

定理6对于“铺盖”凸集的问题是很有意义的。在平面中

K的不重叠全等拷贝的分布称为铺盖。基本问题是：平面的多大部分可以被K的不重叠拷贝所覆盖？换句话说，铺盖所能达到的最高“密度”是什么？如果Q_0是包含K的最小四边形，用Q_0的不重叠拷贝就有可能覆盖平面.然后我们得到用K的拷贝密度$\geqslant\sqrt{2}/2>0.707$的铺盖。这个铺盖甚至有着某种程度的正则性。那就是用K铺盖的两个“点阵”的和集.参考文献[3]包含了大量的关于铺盖问题的有价值的资料。

6. 某些熟知的极值问题

微积分课程中另一个典型练习是：

给定一个半轴为a和b的椭圆．求内接于E有最大面积的矩形R_0。

在解这个问题中，通常假设矩形的边平行于椭圆的轴。为了证明这个假设是合理的，需要知道任何内接于E的矩形R必有它的边平行于轴（假设E不是圆）．让我们看一下如何用仿射变换证明这一事实。

假设R是内接于E的矩形．仿射地变换E到一个圆E^*。那么，在同一个变换下，R变到内接于E^*的平行四边形R^*。证明任何内接于圆的平行四边形必是矩形，这是一个极普通的练习。然而对我们有意义的事实是R^*的中心与E^*的中心重合；因此，R的中心也必与E的中心重合。于是，R的外接圆C也定圆心在E的中心．很显然这样一个圆C与E相交于四个点，这四个点正是其边平行于轴的矩形的顶点；因此，R就是这样一个矩形。

练习4 用仿射变换将求R_0的问题简化为求内接于圆的最大矩形的问题。

下面练习的结果可以用一个仿射变换很容易确立，这个练习与练习2是密切相关的。

练习5 证明椭圆E内部任何三角形的最大面积是$(3\sqrt{3}/4\pi)$·面积(E)。

下面的在[3,p.36]中证明的定理对于我们的关于有最小面积的外切n边形的考察是一个补充：

定理7 如果P是在一个凸区域K内有最大面积的内接n边形，那么

$$\text{面积}(P)\geqslant\frac{n}{2\pi}\sin\frac{2\pi}{n}\cdot\text{面积}(K),$$

而等式仅当K是椭圆时成立。

注释 这篇文章的某些例子在 Klamkin和Newman([6])的论文中也讨论了。下面再给出一个有趣的练习。

练习6 通过椭圆内一给定点，画一条截去最小面积的直线。

虽然还有其它的或多或少是熟知的例子，我们也就此停笔了。

参 考 文 献

[1] M. T. Bird, Maximum rectangle inscribed in a triangle, to appear.

[2] L. Danzer, D. Laugwitz, and H. Lenz, Über das Löwnersche Ellipsoid und sein Analogen unter den einem Eikörper einbeschriebenen Ellipsoiden, *Arch. Math.*, 8 (1957) 214—219.

[3] L. Fejes Tóth, Lagerungen in der Ebene, auf der Kugel, und in Raum, Berlin, 1953.

[4] C. M. Fulton and S. K. Stein, Parallelograms inscribed in convex curves, *Amer. Math. Monthly*, 67 (1960) 257—258.

[5] W. Gross, Über affine Geometrie XIII: Eine Minimumeigenschaft der Ellipse und des Ellipsoids, *Leipziger Berichte*, 70 (1918) 38—54.

[6] M.S. Klamkin and D. J. Newman, The philosophy and applications of transform theory, *SIAM Rev.*, 3(1961) 10—36.

[7] L.H.Lange, Some inequality problems, *The Math. Teacher*, 56 (1963) 490—494.

[8] ——, Elementary Linear Algebra, Wiley, New York, 1968, pp. 140—147.

[9] C. Radziszewski, Sur un problème extrémal relatif aux figures inscrites dans les figures convexes, *C. R. Acad. Sci.* Paris, 235 (1952), 771—773.

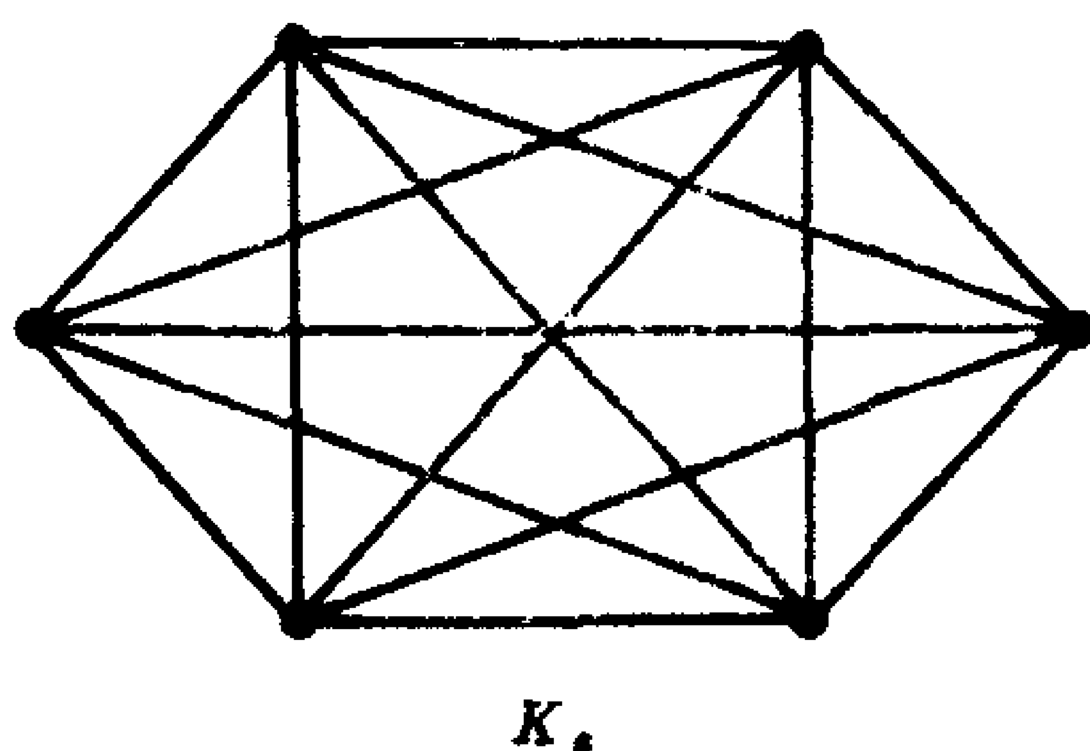
（蒋定华译，刘　勇校）

恰有两个单色三角形的相识图

Frank Harary

Goodman[3]证明了在任何一个有六个人的聚会上，如果任意两个人要么互相认识，要么互不认识，那么不仅存在三个人彼此认识或互不认识（我们已在[1]中知道这个结论），而且至少存在两个这样的三人组.

我们用图论（见[4]）的语言来描述处理这样的问题是非常方便的，本文将引用它的术语及符号。给定 n 个点，并在任意两点间用一条线相联。这样得到的点-线图称为 n 阶完全图，记作 K_n。图中的点及线分别称为顶点及边。例如，6 阶完全图 K_6 有 6 个顶点和 15 条边（见图1）。



K_6

图 1

完全图 K_n 的一个 2-边着色是指将它的每条边涂染为绿

色或红色。如果我们用实线及（长划的）虚线分别表示绿色边及红色边，则图 2 给出了一个使 K_6 含有两个不相交的绿三角形①，但不含红三角形的 2-边着色。如果我们将实线及（长划的）虚线分别看作是正边及负边，则我们得到一个指定符号的图，[5]的第十一章研究了这类图。

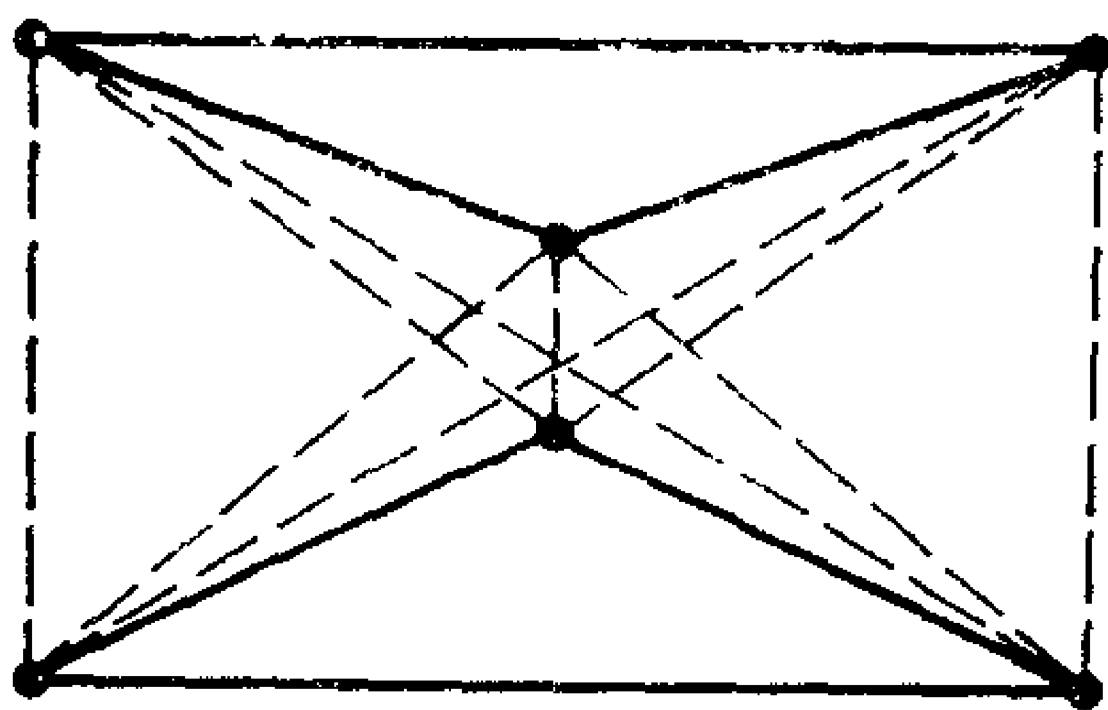


图 2

显然，当一个完全图 K_6 的所有边同色（比如说均为绿色）时，则我们得到单色三角形的最大可能的数目为 $\binom{6}{3}=20$。然而，要确定哪些着色方案使得图 K_6 中恰有两个单色三角形，这并非是一件平凡的事情。本文的目的就是要完全地确定出这些方案。

我们所知道的关于 K_6 的每一个 2-边着色均至少含有一个单色三角形的最简短的证明如下。每个顶点 u 必定通过三条同色边与另外三个顶点 u_1, u_2，及 u_3 连接，比如通过绿色边（如图 3(a)所示）。现考察 u 的这三个相邻点。假若存在某条边 u_iu_j 为绿色，则我们有一个绿三角形 uu_iu_j。否则，$u_1u_2u_3$ 构成一个红三角形，如图3（b）所示。

① 边为同色的三角形称为单色三角形。绿（红）三角形则是指边均是绿（红）色的三角形。——译注

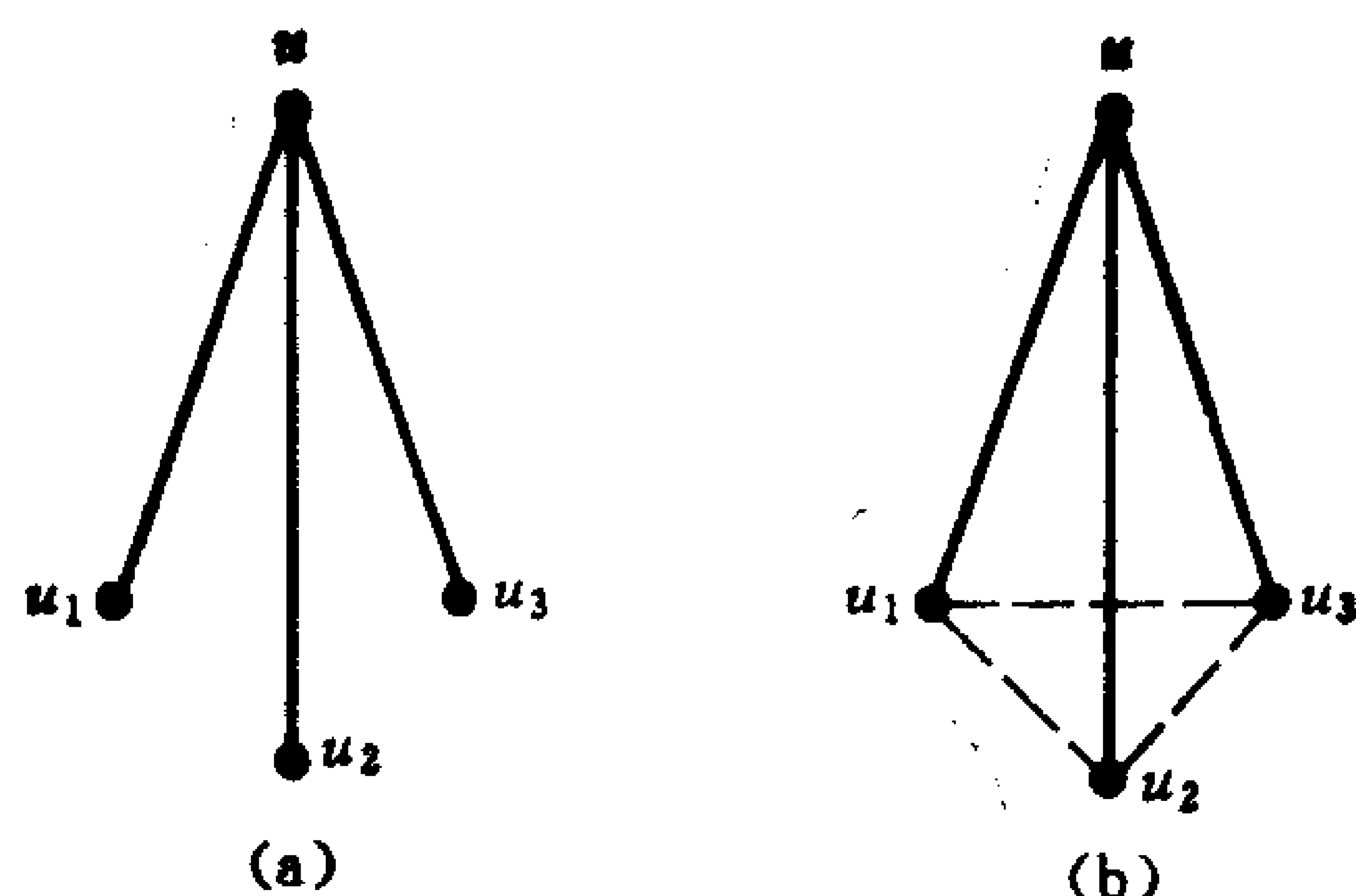


图 3

如果把“顶点”看作是“人”，并用“红边”联结表示“彼此认识”，用“绿边”联结表示“互不认识”，这样的着色图称为相识图，则 Goodman 的结论可表述为：

Goodman 定理（[3]） 完全图 K_6 的每个2-边着色均至少含有两个单色三角形。

下面我们就在 Goodman 定理的基础上，求出这种相识图 K_6 的所有正好含有两个单色三角形 T_1 及 T_2 的 2-边着色方案。很明显，T_1 及 T_2 可以在 0，1 或 2 个顶点上相交。我们将按各种情形进行讨论。

情形0 T_1 与 T_2 没有公共顶点。

这时有两种可能性：

情形0.1 T_1 与 T_2 不同色。

设 $u_1u_2u_3$ 为一个绿三角形，且 $v_1v_2v_3$ 为一个红三角形。不妨假定 u_1v_1 为绿的，如图 4 所示。则边 u_3v_1 必定是红色的，因为否则 $u_1u_3v_1$ 将是第三个单色三角形。类似地，边 u_3v_3 不得不为绿的，且 u_2v_3 只能是红边，如图 5 所示。但是，现在边 u_2v_1 既不能是绿的（因为边

$u_1v_1$① 及 u_1u_2 均是绿的）也不能是红的（由于有红边 u_2v_3 及 v_1v_3），这表明情形 0.1 是不可能发生的。

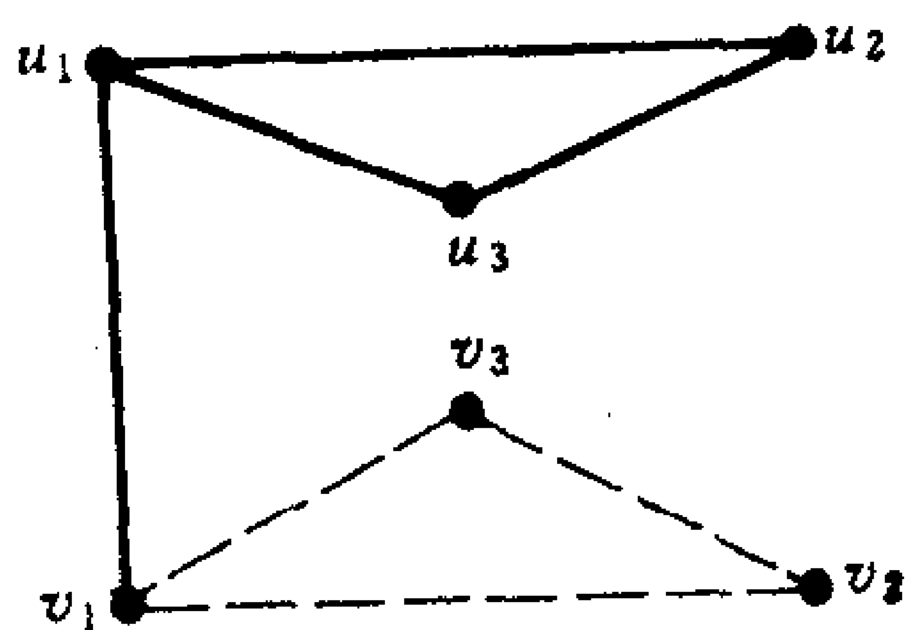


图 4

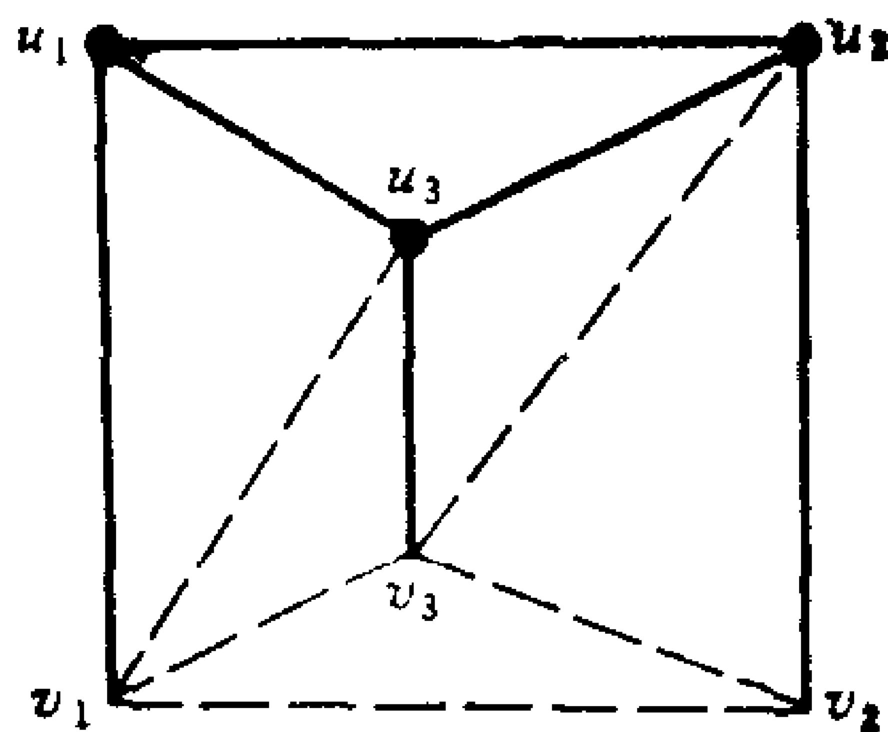


图 5

情形0.2　T_1 与 T_2 同色。

假定 $T_1 = u_1u_2u_3$，$T_2 = v_1v_2v_3$，且它们均为绿的。那么，边 u_1v_1 可以是绿的，但其余边 u_1v_j 及 u_iv_1 均不能为绿的，因为否则我们将得到第三个单色三角形。同样地，边 u_2v_2 及 u_3v_3 可以是绿的。

总之，当 K_6 的一个 2-边着色正好含有两个无公共顶点的单色三角形 $T_1 = u_1u_2u_3$ 及 $T_2 = v_1v_2v_3$ 时，这两个三角形必同色，比如说都是绿色的，则 $u_iv_i (i = 1,2,3)$ 这三条边可以是绿的也可以是红的，但所有其它边都必须为红的。

图 2 给出了 K_6 的一个 2-边着色，它使得 T_1 及 T_2 均为绿的，且其余边均是红的；而图 6 所表示的是 T_1 及 T_2 均为绿的，且 $u_iv_i (i = 1,2,3)$ 这三条边也为绿的。

情形 1　T_1 与 T_2 恰有一个公共顶点。

在这情况下，我们将发现仅存在 K_6 的一个 2-边着色使得 T_1 与 T_2 着有不同的颜色。

① 原文误为 u_1v_2。——译注

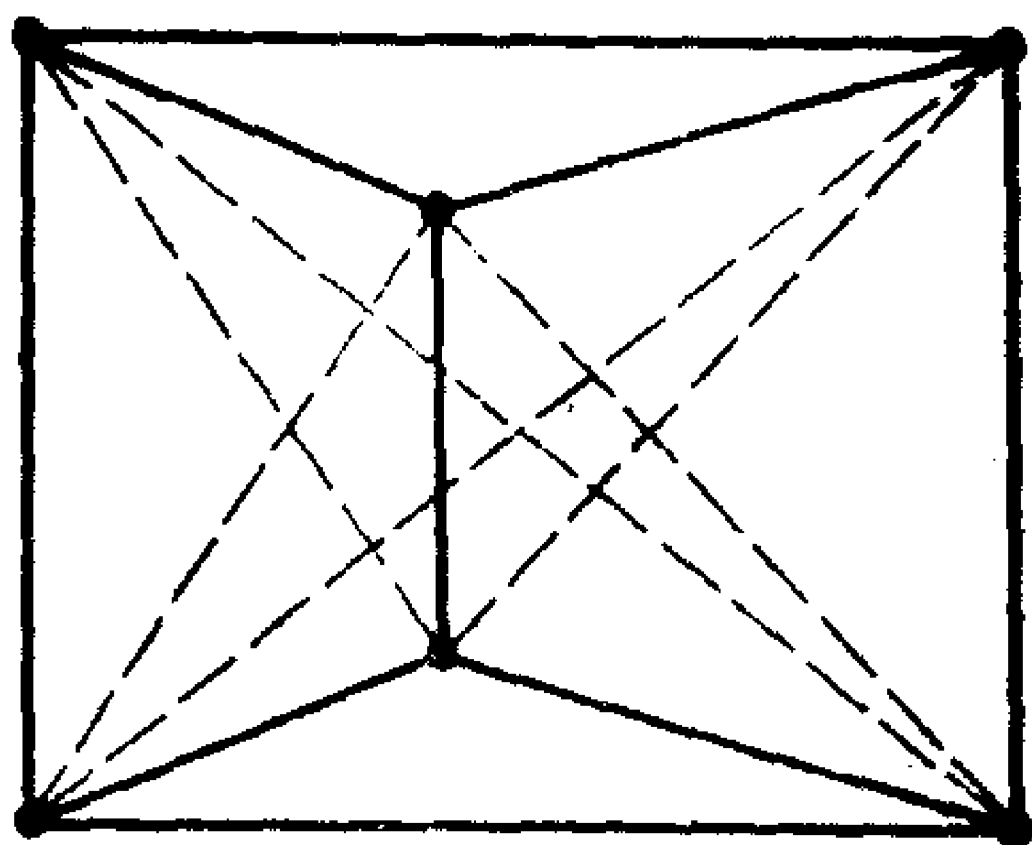


图 6

情形 1.1　T_1与T_2同色。

图 7 (a) 给出了两个绿三角形 $T_1 = uu_1u_2$ 和$T_2 = uv_1v_2$。为了避免出现第三个绿三角形，这五个顶点上的其余四条边均必须是红的，如图 7 (b) 所示。现在考虑第六个顶点 w。假定 wv_1 是红的。则 wu_1 及wu_2 必为绿的，出现了第三个绿三角形 wu_1u_2。因此，边 wv_1 必定是绿的，但 wu 及wv_2 两者均必为红的，由此依次迫使wu_1 及wu_2 均是绿的，如图 7 (c) 所示。可是，这时又出现了第三个绿三角形 wu_1u_2。所以，情形 1.1 是不可能的。

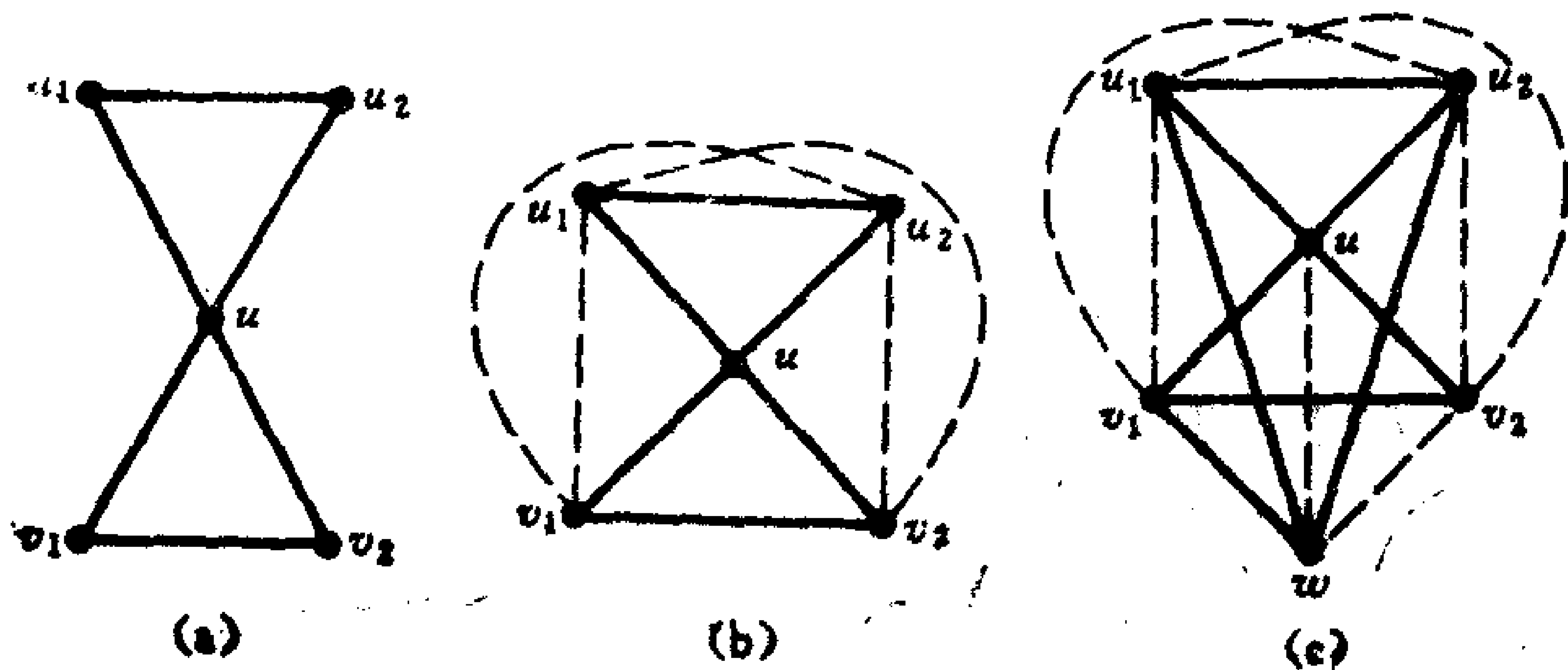


图 7

情形1.2　T_1与T_2不同色。

设$T_1=uu_1u_2$是一个红三角形，且$T_2=uv_1v_2$是一个绿三角形，如图8(a)所示。由于边u_1v_1为何种颜色是无关紧要的，我们可假设它是绿的。如图8(b)所示，我们接连地导致边u_1v_2, u_2v_2，及u_2v_1的颜色必须分别为红的，绿的，及红的。这样，我们完成了对这五个顶点上的十条边的着色，且发现每种颜色的边恰有五条。

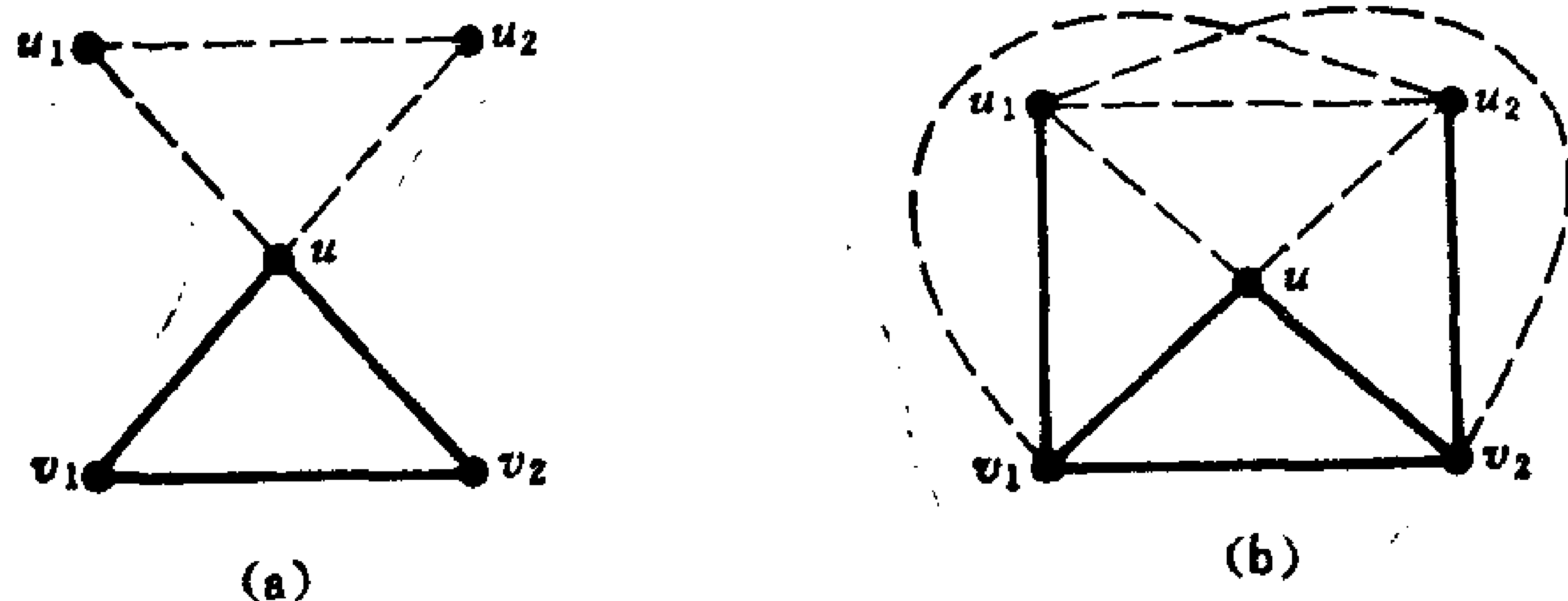


图 8

现在考虑第六个顶点w。若假定wv_1是绿的，则wu及wu_1两者均必为红的。于是wuu_1是一个红三角形。因此，wv_1必须是红的。为避免第三个单色三角形的出现，我们不得不相继地让wu_2着绿色，wv_2着红色，且wu_1着绿色。实际上，到了这步不管让所剩下的边wu着什么颜色都没关系，我们都将仍然仅有原来的那两个单色三角形T_1和T_2。我们在图9中用打点的虚线来表示边wu的这种自由选择。

情形2　T_1与T_2有一条公共边。

显然，当两个三角形有公共边时，它们必着有同一颜色。不失一般性，我们假设T_1与T_2均是绿的。对于情形2中的各种可能的细节，我们可以完全类似于情形0及情形1

来进行分析讨论，最后得到 K_6 的唯一的 2-边着色，如图10所示。

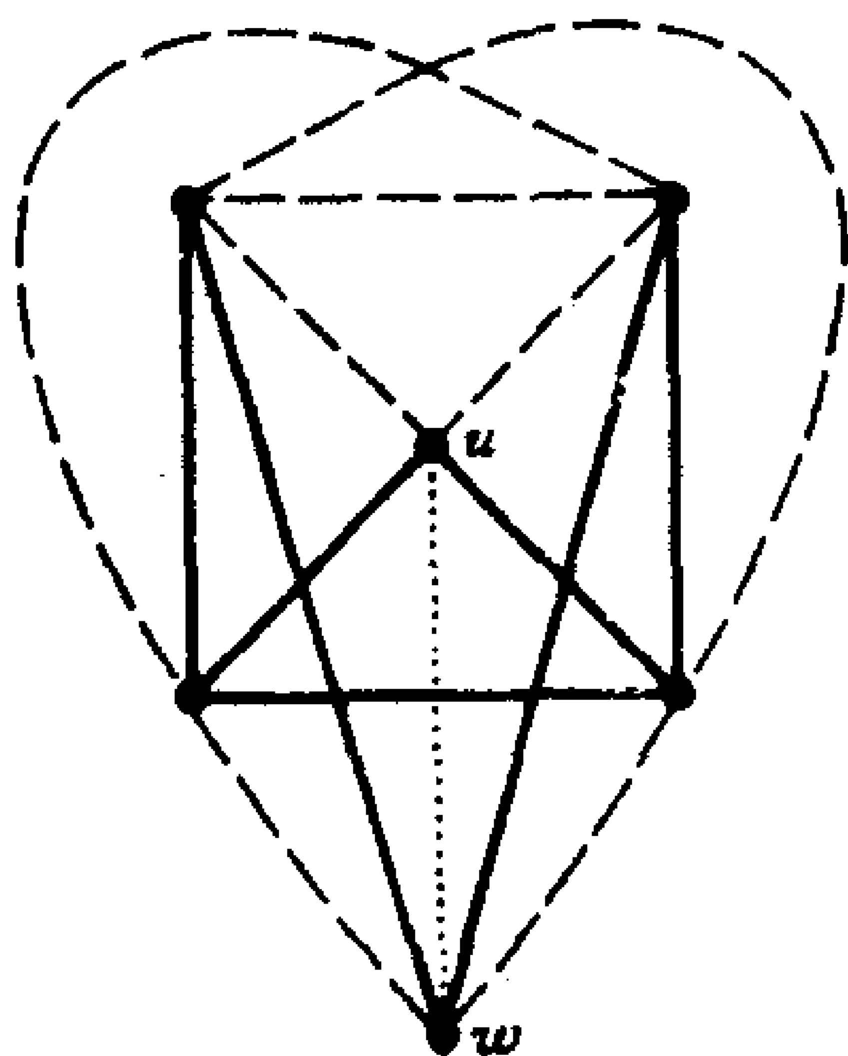


图 9

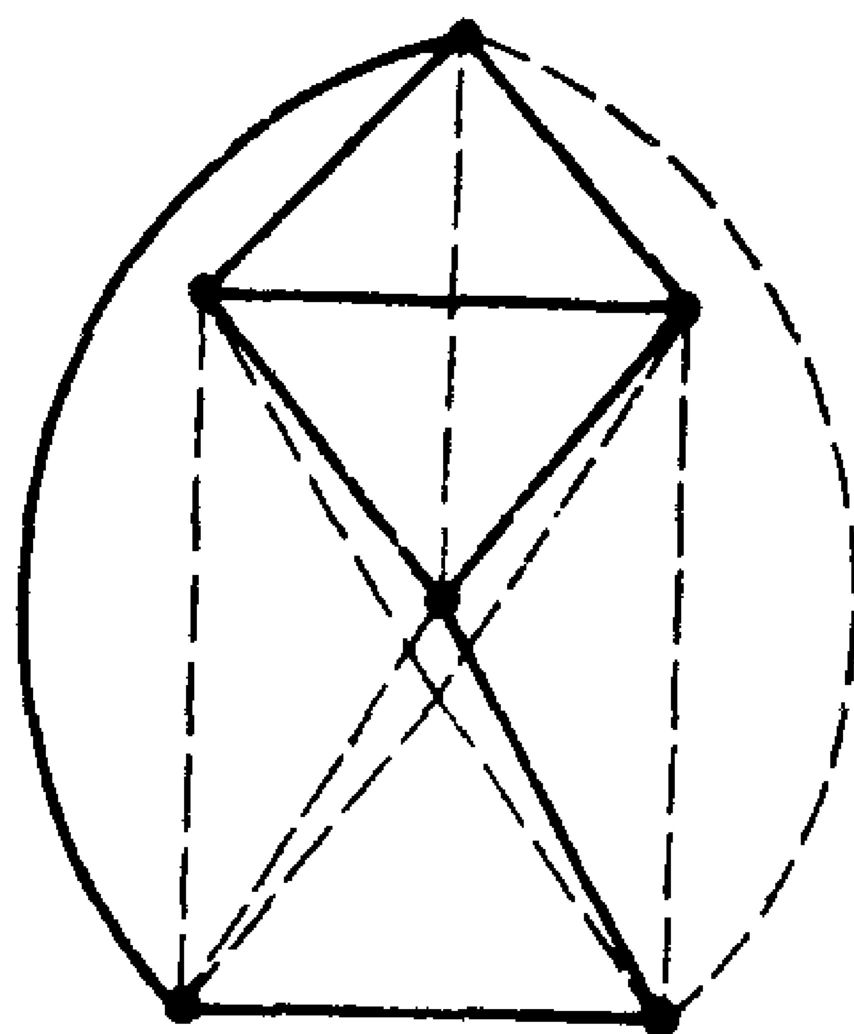


图 10

综合上述各种情形，我们得到

定理 存在 K_6 的 2-边着色使得它正好含有两个单色三角形 T_1 及 T_2，这里 T_1 与 T_2 可以有 0,1 或 2 个公共顶点。而且，T_1 与 T_2 不同当且仅当它们仅有一个公共顶点。

K_6 的所有这样的 2-边着色分别由图 2,6，9 及10所示。

在我们关于图的 Ramsey 推广理论的系列文章[2]的概念中包含了上述问题。而且许多其它的相似问题可如下产生。给定一个无孤立点的其它小子图 F①，试确定完全图 K_n 的哪些2-边着色所含单色子图 F 的数目正好达到最小。

A.Schwenk 在阅读本文的初稿时，从上列图中注意到在 K_6 的每一个含有最少（即 2 个）单色三角形的 2-边着色中，那两个单色子图上的每个顶点的度② 几乎相等。并且，他成

① 如取四边形，五边形等。——译注

② 图的顶点的度是指图中与它相联的边的数目。——译注

功地将这一观察到的结果推广到任何完全图 K_p 的含有最少单色三角形的2-边着色上。而这个最少单色三角形的数目已由 Goodman 精确地给出：

定理（[3]） 设 t 是 K_p 的一个 2-边着色中所含单色三角形的数目，则

$$t \geqslant \binom{p}{3} - \left[\frac{p}{2}\left[\left(\frac{p-1}{2}\right)^2\right]\right].$$

下面的定理已蕴含在 Goodman 的论文中，Schwenk[6] 依靠这个定理不仅得到了对 Goodman 的结果的另一证明，而且作为定理的推论，他获得了一种不需要上述那些令人筋疲力尽的推理过程就能导出图 2, 6, 9 及 10 中所表示的 K_6 的2-边着色的方法。

像书[4]一样，我们用 $\delta(G)$ 及 $\Delta(G)$ 分别表示图 G 的顶点的最小度及最大度。而且，对实数 x，我们记不小于 x 的最小整数为 $\{x\} = -[-x]$。则我们有

定理（[6]） 上面定理中的 t 达到下界当且仅当在 K_p 的一个2-边着色中的每个单色子图 G 的所有顶点的度均尽可能地接近 $(p-1)/2$，使得当 $p \not\equiv 3 \pmod 4$ 时，

$$\left[\frac{p-1}{2}\right] \leqslant \delta(G) \leqslant \Delta(G) \leqslant \left\{\frac{p-1}{2}\right\};$$

而且当 $p \equiv 3 \pmod 4$ 时，G 恰有一个顶点的度为 $(p-3)/2$ 或 $(p+1)/2$，而其余顶点的度均为 $(p-1)/2$。

参 考 文 献

[1] C. W. Bostwick, E 1321, *Amer. Math. Monthly*, 66(1959), 141—142.

[2] V.Chvátal and F.Harary, Generalized Ramsey theory for graphs I,II,III, to appear.

[3] A.W.Goodman, On sets of acquaintances and strangers at any party, *Amer.Math.Monthly*, 66(1959), 778—783.

[4] F.Harary, Graph Theory, Addison-Wesley, Reading, 1969.

[5] ——, R.Z.Norman and D.Cartwright, Structural Models, An Introduction to the Theory of Directed Graphs, Wiley, New York, 1965.

[6] A.Schwenk, The acquaintance graph revisited, *Amer. Math. Monthly*, to appcar.

（陈赐平译，潘承彪校）

纽结理论中的新型不变量①

L.H.Kauffman 原著，王幼宇改编

编者按 Jones 在1984年发现了一个新的纽结不变量，因而获得了1990年京都国际数学家大会上颁发的Fields奖。后来，Kauffman发现了Jones 多项式的一个完全初等的讲法，本文就是他的这种初等讲法的介绍。详细的内容可以参看 L.H.Kauffman, On Knots, Princeton University Press, 1987, Princeton, New Jersey.

我们希望通过本文使读者对这类新型纽结不变量有所了解，并且产生兴趣。

§1 导 引

1.1 从三叶结（trefoit）谈起

三叶结（见图1）是一个典型的纽结（knot）；它可以看成由一条绳子打一个结后，接上两端而形成的（如图2所示）。如果我们不打结，而直接将绳子两端接上，得到的就是普通的圆圈（unknotted circle）。

---

① *New Invariants in the Theory of Knots, Amer. Monthly of Mathematic*, 95(1988), 195—242. 王幼宇根据其中的部分内容重新编写，并且作了许多解释、*Kauffman* 的原文还包括其它类型的不变多项式及其相互关系，此外还有纽结理论在图论和统计物理学中的应用。读者可进一步参考原文.

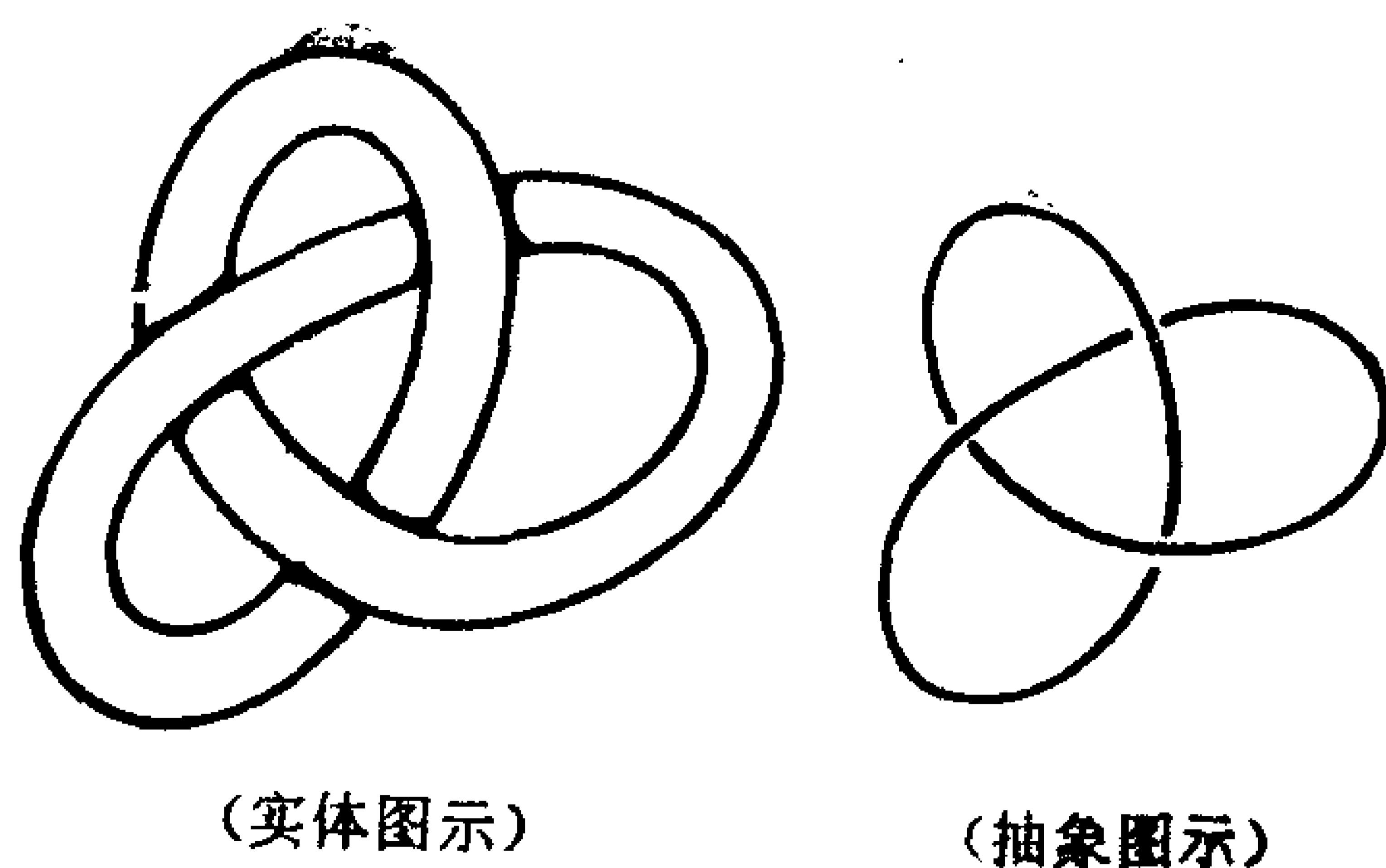


图 1

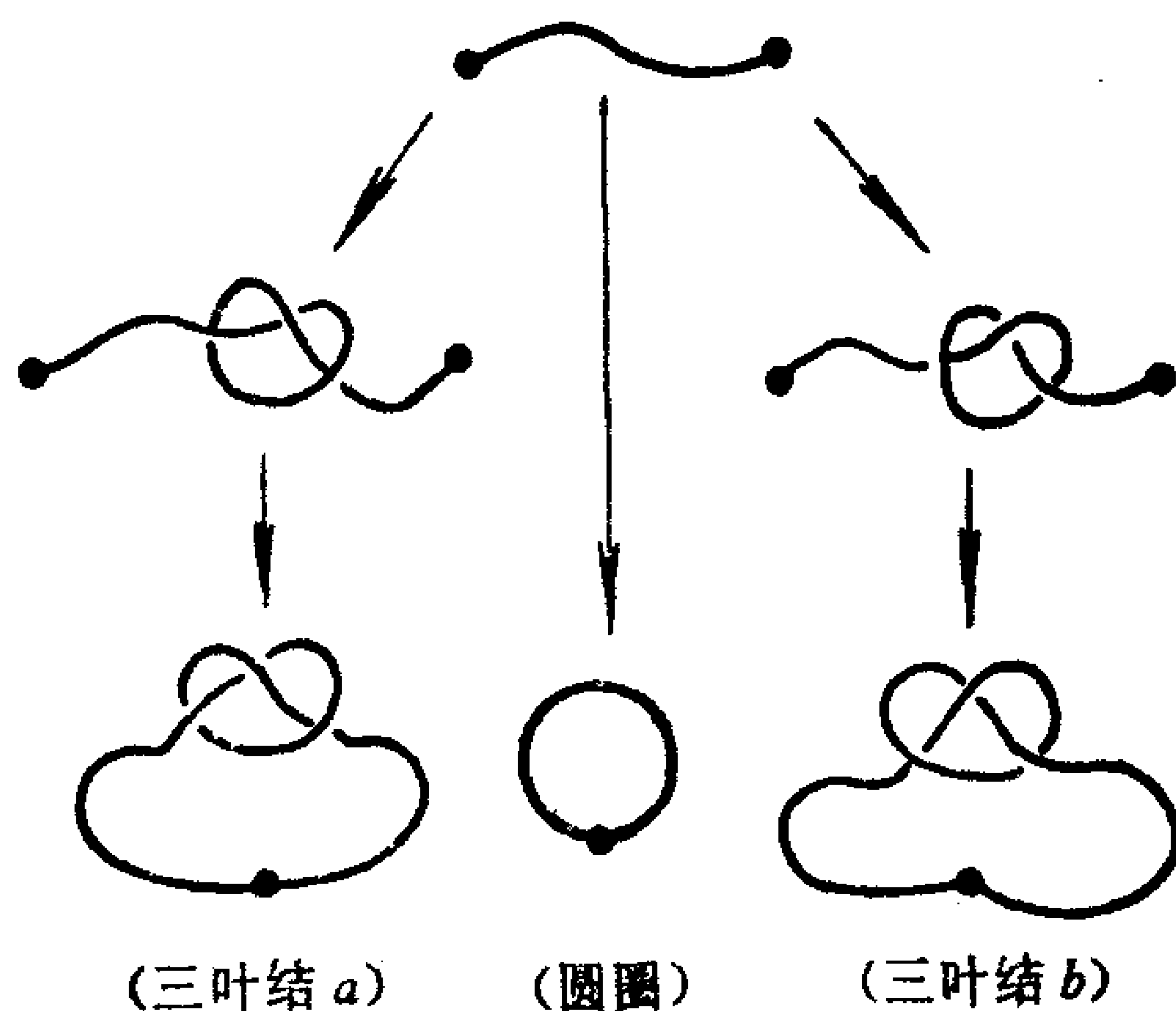


图 2

直感可以告诉我们，三叶结与普通的圆圈在拓扑上是不相同的。也就是说，无论怎样变形，只要不使绳子断开，我们就不可能把它们的形状变成一样的。

怎样说明我们的直感的正确性呢？我们可以用它们所附带的、具有拓扑不变性的量，即通常所说的拓扑不变量，把

它们区分开。本节第四段，将说明三叶结确实是一个纽结（区别于普通的圈）。在§2中，我们还将进一步证明三叶结是有旋向(chiral)的，即它自身与其镜像（mirror image)具有不同的拓扑。也就是说，图2中所给出的 a,b 两种三叶结（互成镜像），在拓扑上有着本质的区别。

1.2 平面图形

在前文中，我们已用平面图形表示三维欧氏空间中的实体，这样做是清晰和有效的。如图2所示。

实际上可以这样来看：我们把纽结或环链(link)“摊开”在平面上，出现在我们面前的，是一些平面曲线段及其交叉(crossing)。这里的“交叉”有两种情形（见图3)，分别对应于曲线段相交时两种不同的“上”、“下”相对位置（参见图1）。同样，最简单的环链如图4所示。

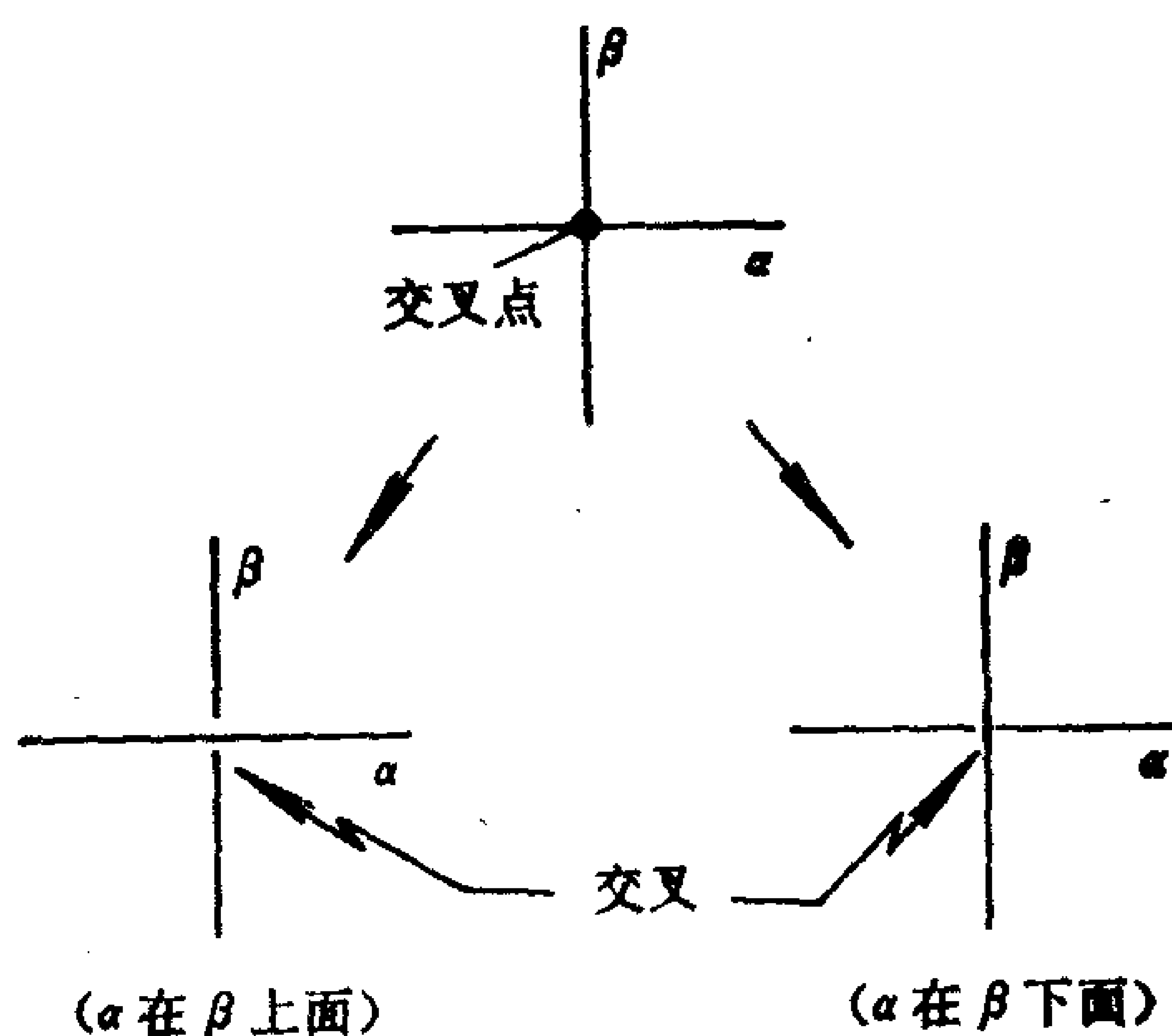


图 3

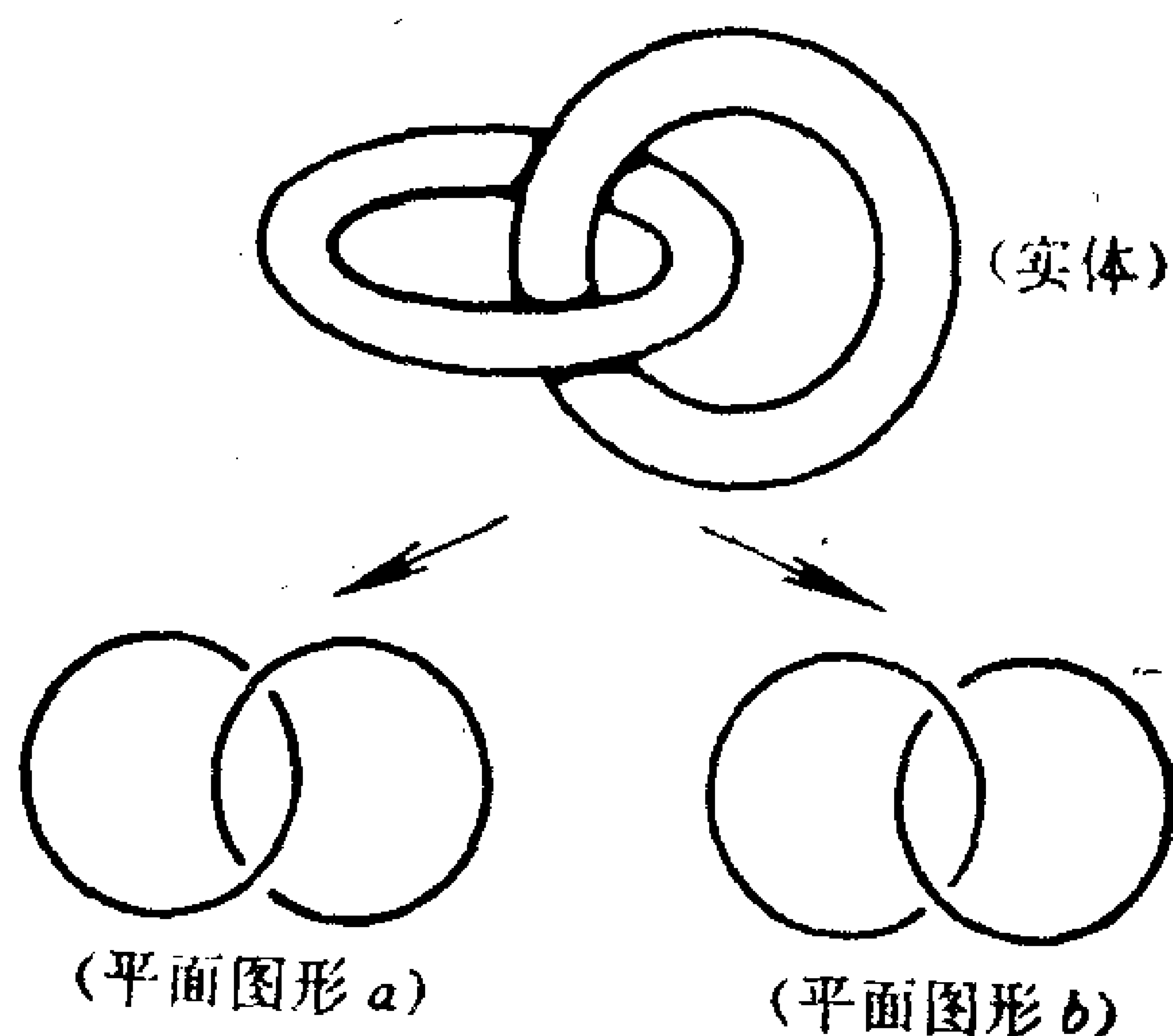


图 4

换言之，我们可以将纽结和环链看成是在一个4-价平面图（即在每个顶点有四个价键的平面图，也就是每个顶点只是两段曲线的交点）上的附加结构（即选择如图 3 所示的一定的“交叉”方式），其中的平面图称为它的通用像（universe）。图 5 画出的是三叶结及其通用像。

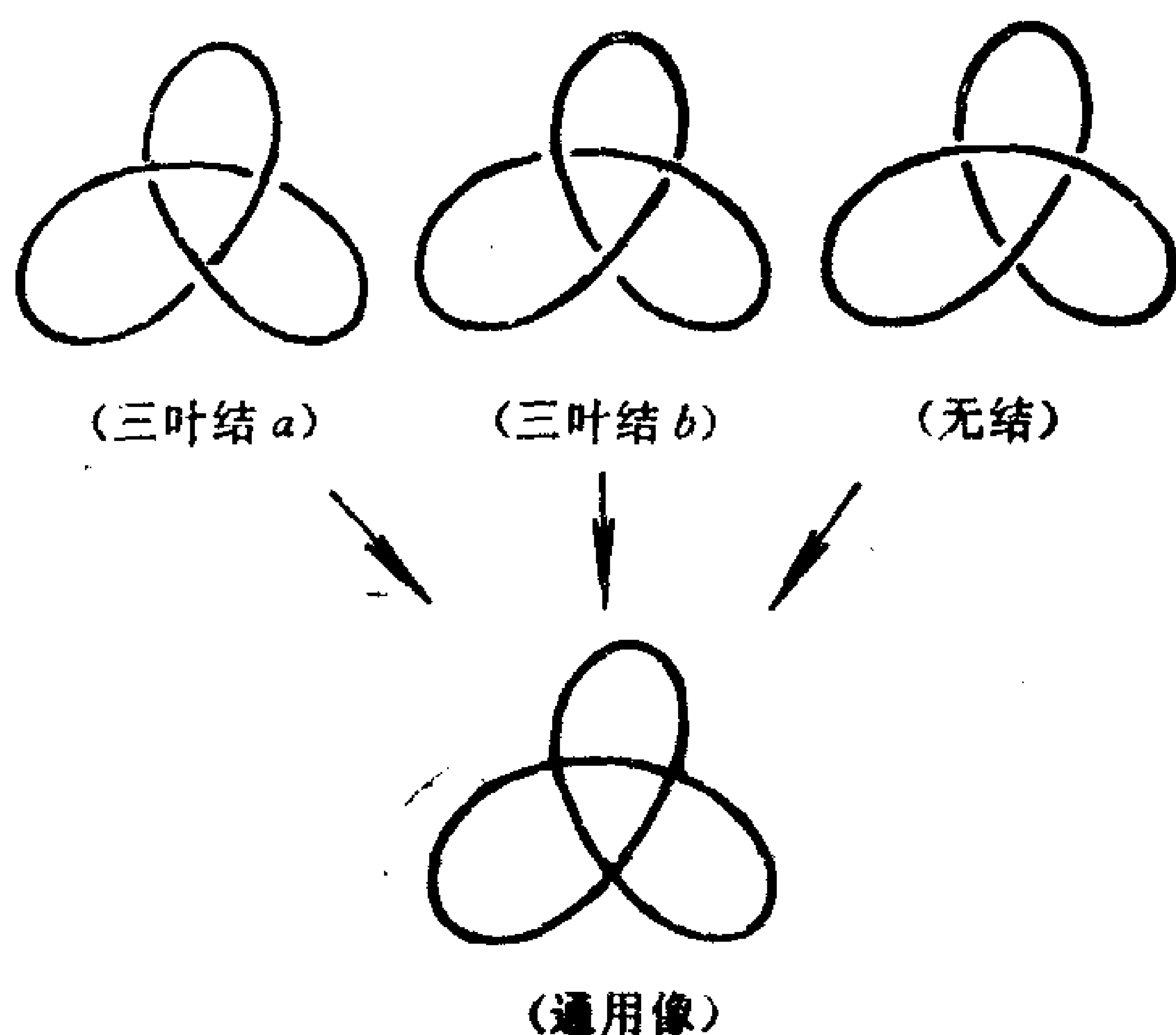


图 5

为便于分析研究起见，以后我们所谈到的组结或环链，都是指它们在平面上的相应图形。但是我们应该定义等价关系及其等价类，使得表示三维空间中同一个组结或环链的所有平面图形是彼此等价的。比如，图 4 显示了双环链的两种平面图形，这两个图形应该是等价的。

1.3 图形的等价

形象地说，通用像就是组结或环链在平面上投射的影子。一般地，一个具有 n 个顶点的通用像，是2^n个对应的组结或环链的投影；其中包括了许多没有结扣的组结和没有环绕的环链（参见图 5）。于是，我们需要定义等价关系的另一目的，是使“无结”和“无环套”具有明确的意义。

组结（或环链）的等价是由以下四种平面图形基本运动来实现的。一种是所谓的组结（或环链）的通用像的变形（严格地说即通用像在平面中的外围空间的同伦变形），并且保持各顶点的交叉结构不变，简称为平面合痕（参见图6）。

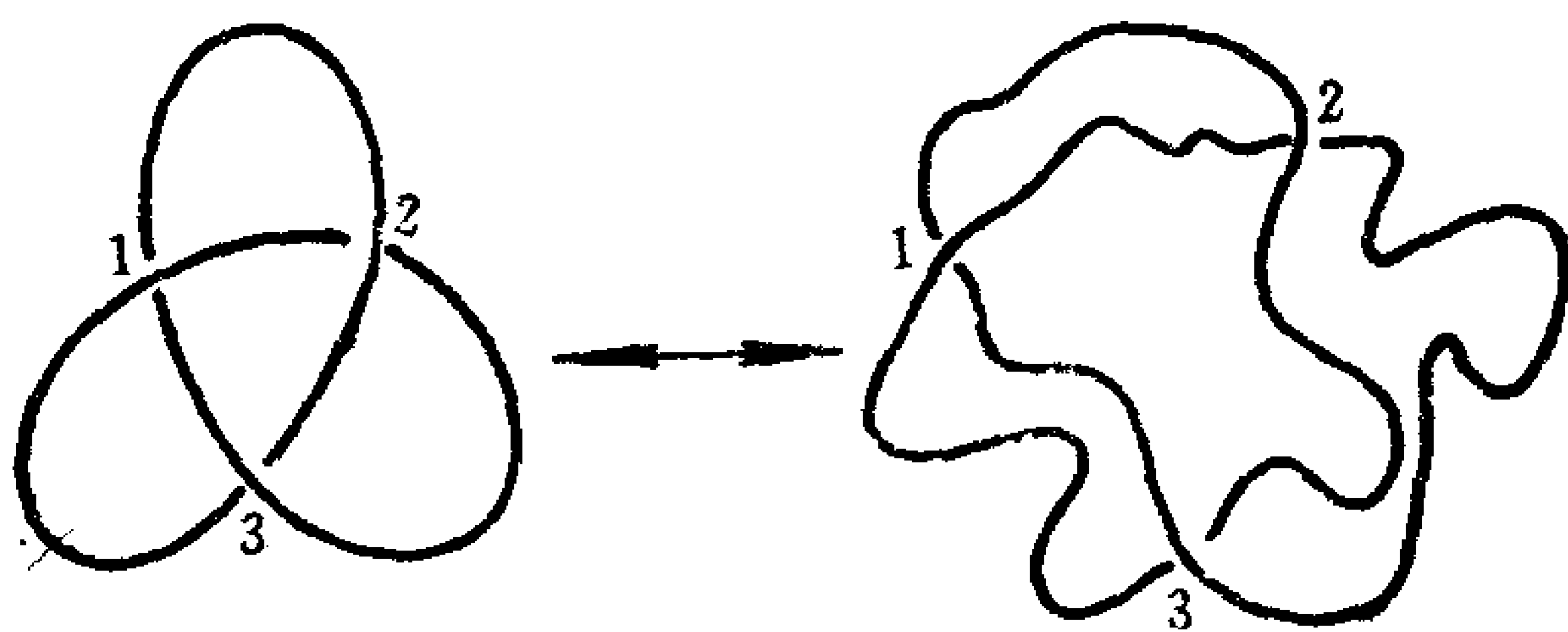


图 6　平面合痕

另外三种基本的图形运动统称为Reidemeister 运动，简称为R-运动，图 7 标明了这三种R-运动：

Ⅰ. 增加或去掉一个曲卷式交叉；

Ⅱ. 移去或增添两个邻接的下（或上）方交叉；

Ⅲ. 三角形运动。

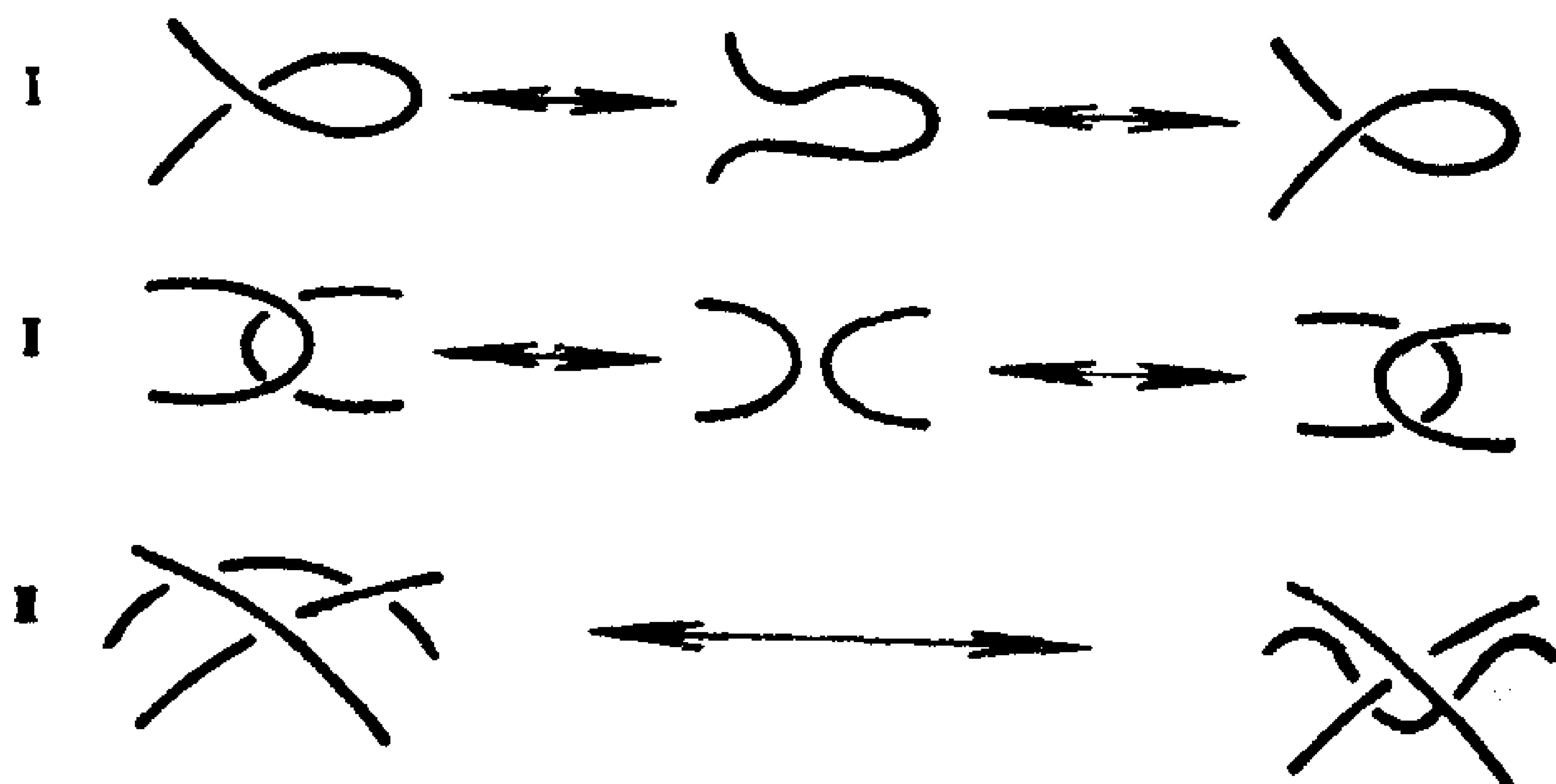


图 7　各种类型的 R-运动

在二十年代，Reidemeister 证明以上四种运动生成了组结（或环链）的空间合痕。也就是说，空间中的两个组结（或环链）可相互变形（在外围空间同伦意义下），当且仅当它们的平面图形可以通过以上四种运动来相互变换。

图 8 给出了变形为无结环的空间合痕。图 9 解释了 8 字形组结 E 与其镜像 E^* 之间的空间合痕，其中最后两步是平面合痕。

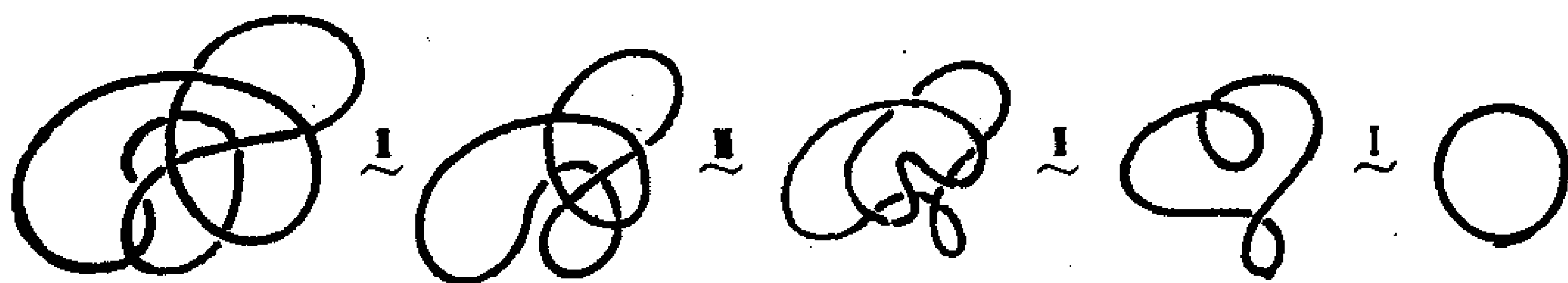


图 8　外围空间合痕

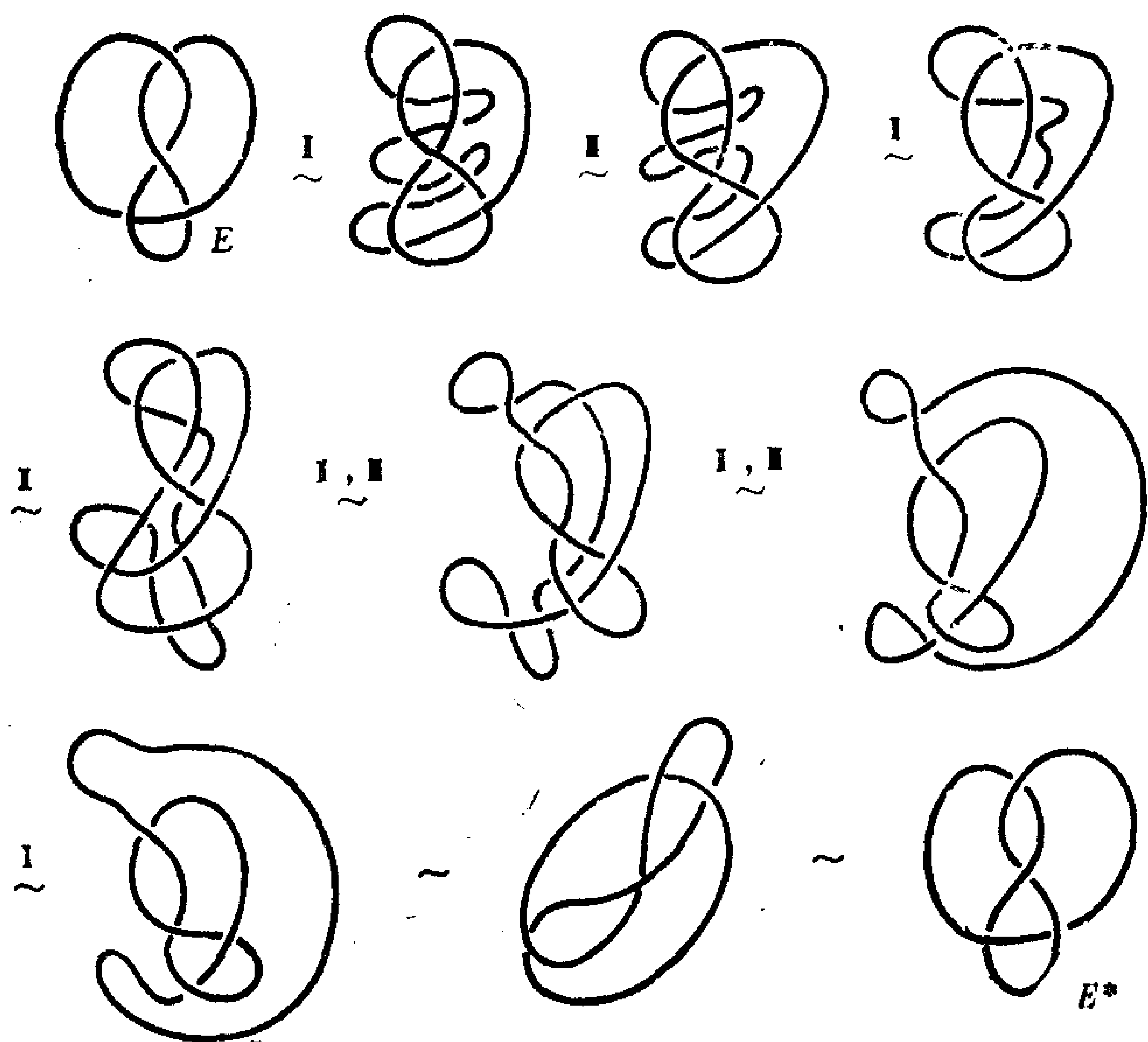


图 9

于是，我们可以定义等价关系及其不变量如下。称两个纽结（或环链）是等价的，如果存在一系列（有限多个）R-运动和平面合痕将其中一个变成另一个。称纽结（或环链）的一个量或一个性质是不变的，如果它在纽结(或环链)的等价变形下是保持不变的。所以我们所说的不变量就是纽结（或环链）的外围空间的拓扑不变量。

纽结（或环链）的分支数显然是一个不变量。这只要注意到，平面合痕和R-运动都不改变分支数。无论图形多么复杂，我们都可以用下述方式来确定它的分支数：任选弧上一点，沿弧走完一圈（途中可能要经过若干交叉点），即确定

出一个分支；再选取不在该分支上的其他弧上的点如法炮制，便可确定下一个分支。参见图10。

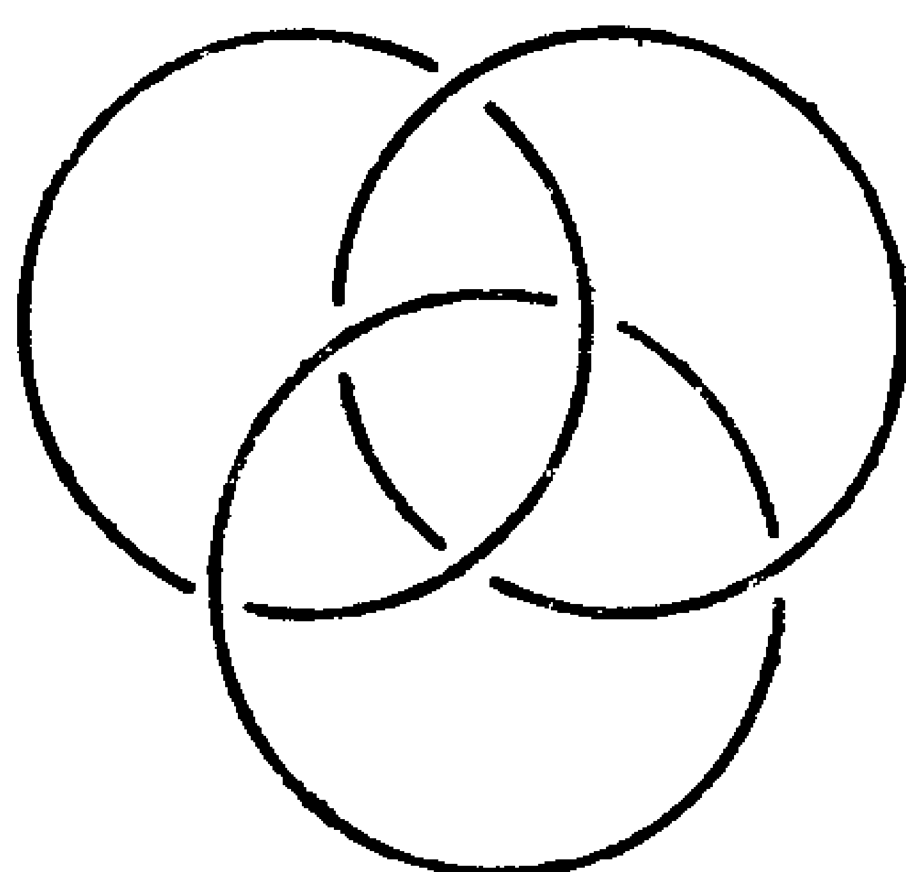


图10　三个分支的环链

由分支数的不变性知道，◯ 和 ○◯ 是不等价的。但仅由分支数，尚无法区分○○和 ⚭ ，也无法区分圆圈与三叶结。这说明分支数是一个很弱的不变量。所以我们需要寻求及研究各种不变量和不变性，以便能够区分不等价的纽结（或环链）。

1.4　三叶结是真纽结

在§2中，我们将利用不变多项式获得更强的结论。在此，我们先作一些最初等的讨论。

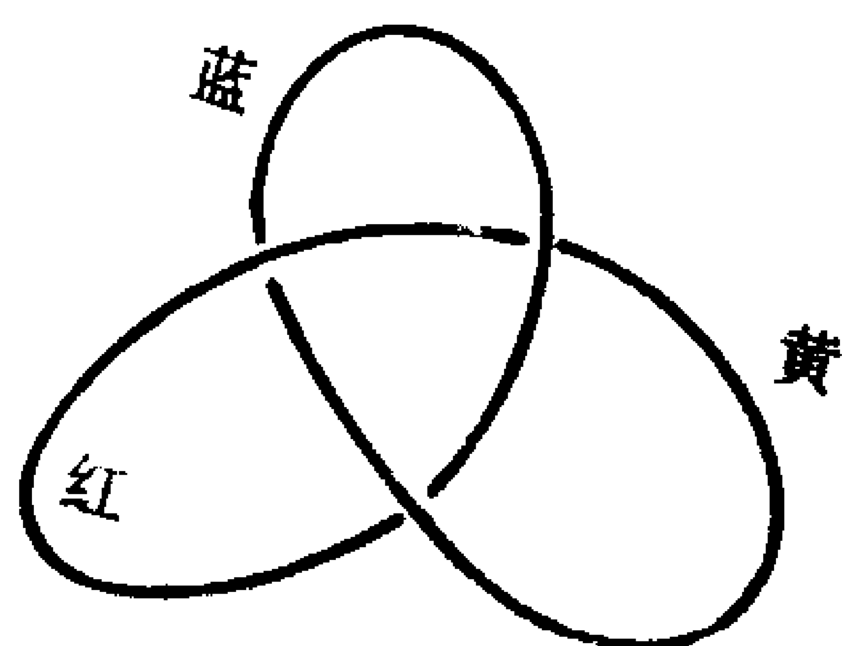


图11　三叶结的三着色

如图11所示，我们在三叶结图形上染上三种颜色——红、黄、蓝。称一个组结图形是可三着色的，如果存在一种染色方式，使

它的每段弧染成红、黄或蓝色中的一种，并且这三种颜色的弧都要有，此外在任何交叉处（看成三条弧 1 2 3）或者出现三种颜色或者只出现一种颜色。显然，三叶结是可三着色的。另一方面可以证明(留作练习)，纽结(一个分支)的可三着色性 在 R-运动下保持不变；从而可三着 色性是纽结的一个不变性。进一步，注意到无结的圈是不可三着色的（它只有一条闭合弧），这就清楚地说明三叶 结 不可能是无结的。

从证明的过程中我们看到，通过对不变性和不变量的分析研究，可以使一些奇妙的结果清晰地显示出来。随着各种各样的不变量的发现和研究，我们就会走进一个丰富多采的世界，在那里充满着朴素而又深刻、独特的思想。

1.5 环绕数 (linking number)

现在我们引进环绕数这个不变量，给两条曲线互相纠缠的程度提供一个衡量的尺度。

一个环链称为定向的，如果它的每个分支都沿着弧标定一个方向，并且这个方向在通过交叉点时不改变。如图12所示，双环链有四种可能的定向。

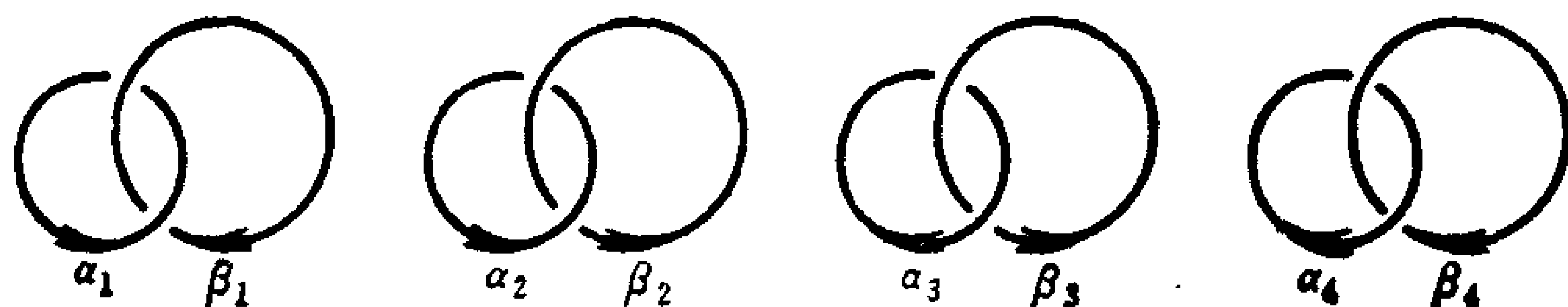


图12 双环链的四种定向



图13 交叉符号和交叉角范围

对定向环链的每个交叉点，我们按图13中所示的方式给出交叉符号（crossing signs），即：设在交叉点处从上方弧段的正方向到下方弧段的正方向的角为θ，则交叉符号是

$$\varepsilon=\operatorname{sign}\theta=\begin{cases}+1, & \text{当}0<\theta<\pi;\\ -1, & \text{当}-\pi<\theta<0.\end{cases}$$

假定一个定向环链的两个分支是α和β，以$\alpha\cap\beta$记α与β的交叉全体（不包括自身交叉），则α与β的环绕数定义为

$$lk(\alpha,\beta)=\frac{1}{2}\sum_{p\in\alpha\cap\beta}\varepsilon(p),$$

其中$\varepsilon(p)$是p的交叉符号。用语言叙述就是：两条定向曲线的环绕数是它们的交叉符号和的一半。例如在图12中，

$$lk(\alpha_1,\beta_1)=\frac{1}{2}(1+1)=1,$$

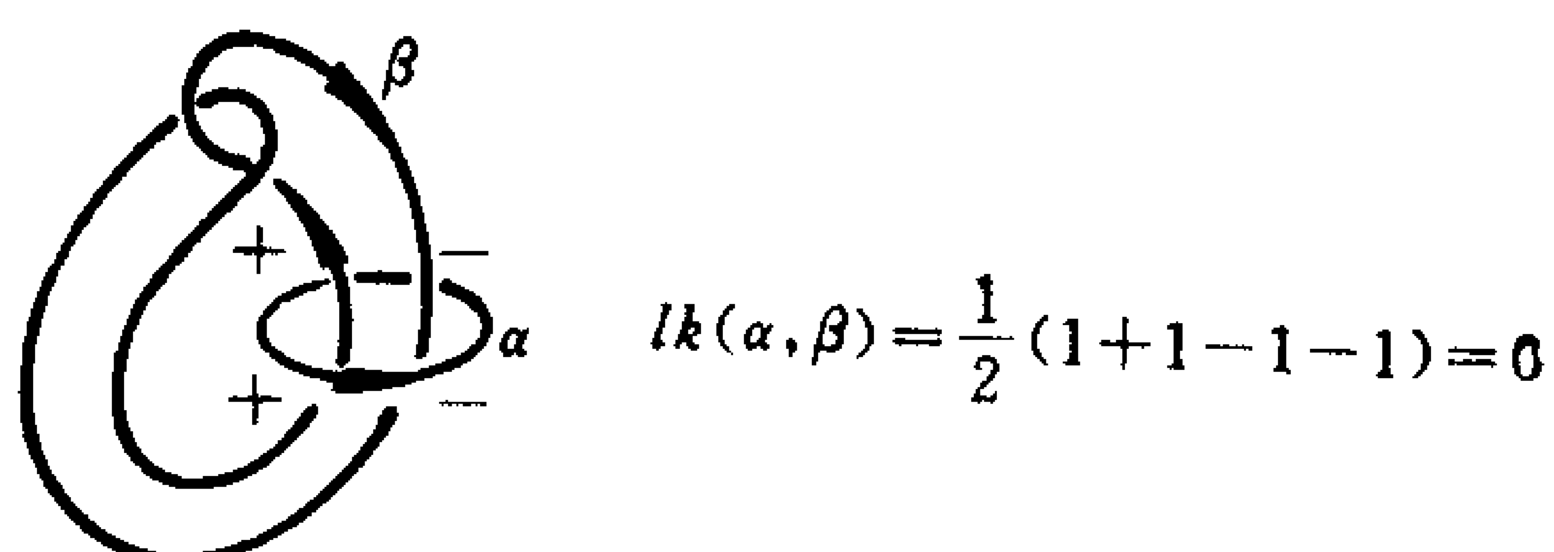
$$lk(\alpha_2,\beta_2)=\frac{1}{2}(-1-1)=-1,$$

$$lk(\alpha_3,\beta_3)=lk(\alpha_1,\beta_1)=1,$$
$$lk(\alpha_4,\beta_4)=lk(\alpha_2,\beta_2)=-1.$$

一旦标定两个分支的方向，直接分析R-运动就得到环绕数是一个不变量（对指定方向而言）。事实上，Ⅰ型运动与环绕数没有关系；Ⅱ型运动同时添加或减少各一个$+1$和-1，从而总和不变；Ⅲ型运动也不改变交叉符号的总和。

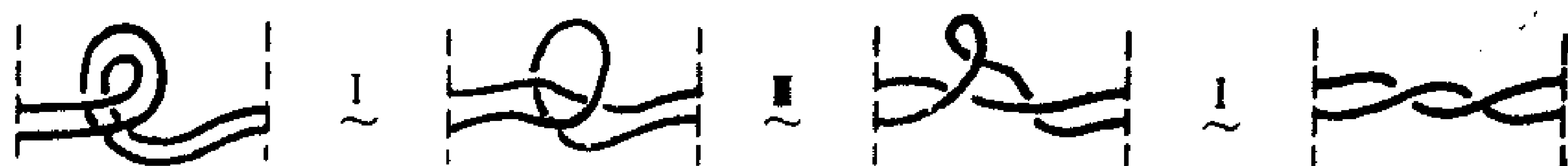
显然，无论如何定向，α β 的环绕数是零。而双环链无论怎样定向，环绕数总是非零的。这足以说明双环链确实是相互环绕在一起的。下面再看一些例子。

例 1



这是以拓扑学家 J.H.C.Whitehead 命名的环链。虽然其环绕数为零，但它是相互绕在一起的。这件事情在§2中将会进一步说明。

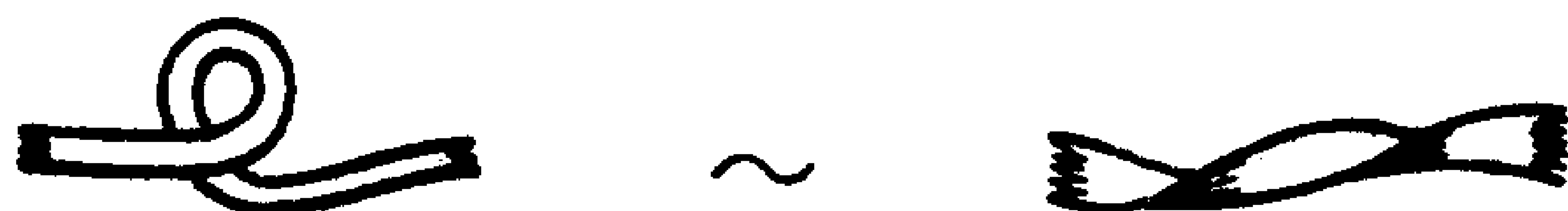
例 2



图示的等价过程表明：保持端点固定，我们有



这个事实可以通过如下演示得到解释：把两条弧看成一条带子的两边，显然有



如果在带子的两条边上标以同样的方向，我们就可以看到，它们的环绕数是一样的：

$\frac{1}{2}(-1-1)=-1$

$\frac{1}{2}(-1-1)=-1$

(自身交叉点不计数)

值得注意的是，卷曲形式 对于环绕数的贡献与自相交叉 的交叉符号是一致的

利用以上的观察，可以求更复杂的环链的环绕数，例如

α β

每个曲卷给出+1

出现两次，共给出$\frac{1}{2}(1+1)=1$

所以

$$lk(\alpha,\beta)=2\times1+1=3.$$

事实上，对于一个纽结图形可以加上一条平行边成为一个双环链，因而我们可以求出所得环链$\hat{K}$的环绕数，它恰好等于原纽结K的交叉符号之和$w(K)$。例 如

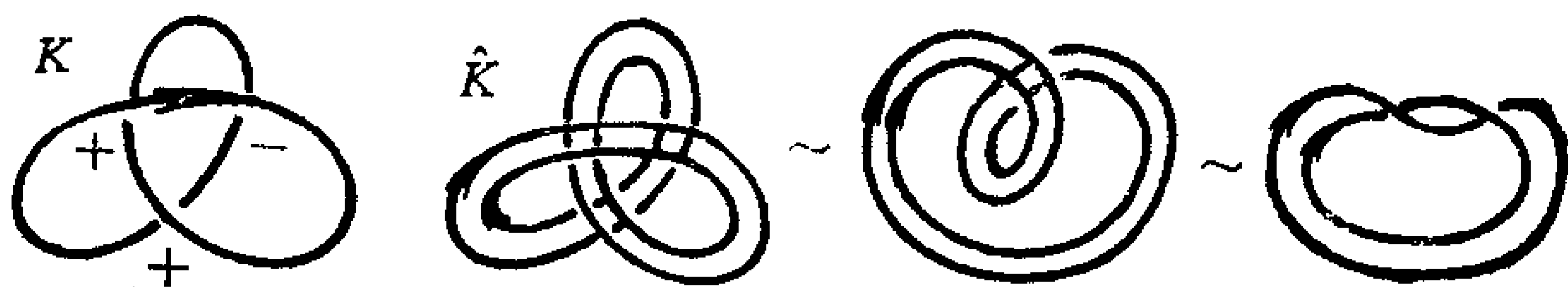
+ + + α β

(纽结 K)　　　　(环链 $\hat{K}$)

$$lk(\alpha,\beta)=w(K)=3.$$

这里的 $w(K)$ 称为 K 的拧数(writhe)，它并不是 K 的不变量，因为在Ⅰ型运动下它的值将改变 ± 1，但它却是如上所述的伴随环链 $\hat{K}$ 的一个不变量。

例 3



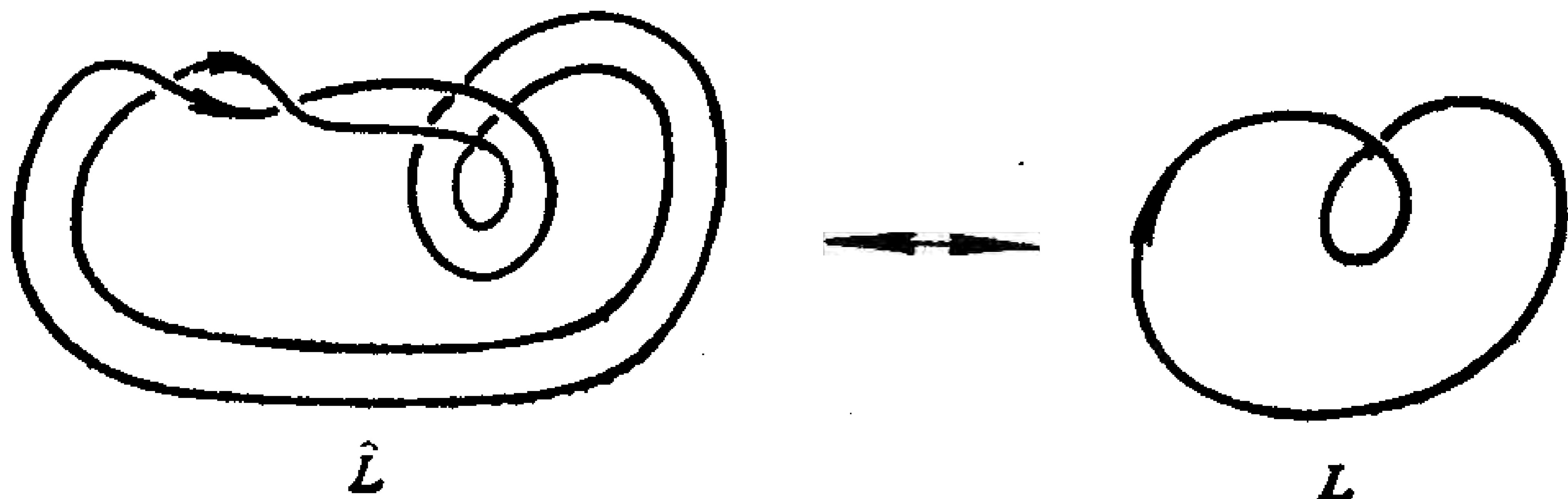
$$lk(\hat{K})=w(K)=+1.$$

对于伴随平行环链 $\hat{K}$ 而言，称 $w(K)$ 为 $\hat{K}$ 的拧数是适当的；我们把盘绕数 (twist number) $T(\hat{K})$ 用于绕在一起的两条边。我们称 为一次完全正盘绕，记作

$T(\quad)=+1$. 于是，对于缠在一起的平行边所组成的环链 $\hat{L}$，我们就有公式

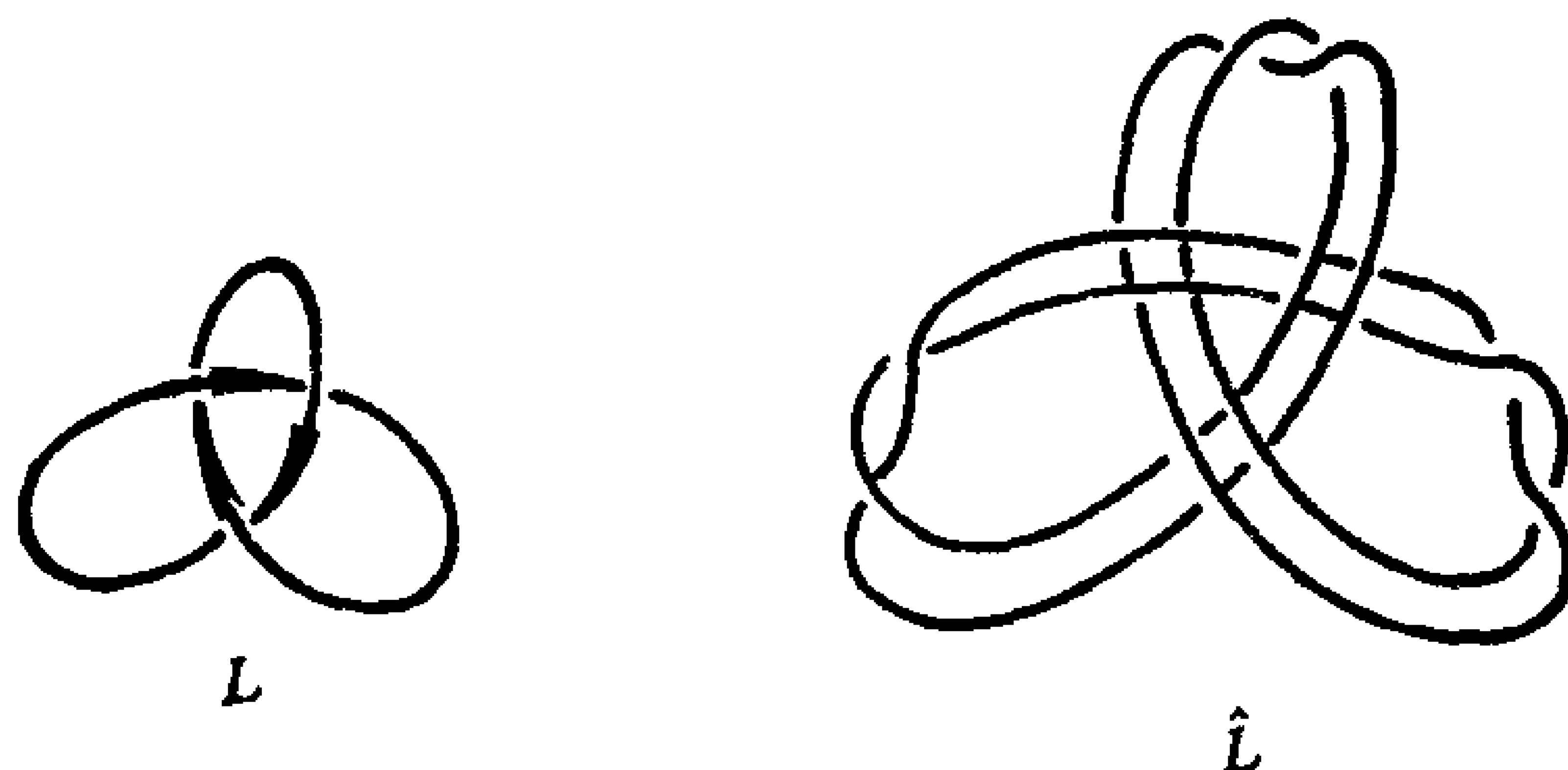
$$lk(\hat{L})=w(L)+T(\hat{L}),$$

即：两条平行的绕在一起的边的环绕数，是拧数和盘绕数之和。例如



$$lk(\hat{L}) = w(L) + T(\hat{L}) = 1 + 1 = 2.$$

例 4



$$w(L) = +3, \quad T(\hat{L}) = -3,$$
$$lk(\hat{L}) = 3 - 3 = 0.$$

这是具有零环绕数却环绕在一起的环链的又一实例。在§2之后，我们可以证明它实际上是有旋向的。

在本节结束之前，应该特别指出的是，公式 $lk(\hat{L}) = w(L) + T(\hat{L})$ 可被认为是闭合平行环链的一种“守恒律”。$w(L)$ 和 $T(\hat{L})$ 每一个都不是拓扑不变量，但由于它们的和是拓扑不变量，从而它们的和一定是常量（在等价关系下）。这个观点已经用来帮助理解双边 DNA 的几何，从而使组结理论与分子生物学之间在七十年代末期产生奇妙的结合。

§2 括号多项式

2.1 方括号多项式

本节的主要目标是引进新型不变量——不变多项式。为此，我们先对非定向环链（或组结）的每个图形定义对应的多项式。

记 $Z[A,B,d]$是变元 A，B和 d 的三元可交换多项式的集合。对给定的非定向图形 K，我们用下述公理来确定相应的多项式$[K]\in Z[A,B,d]$。

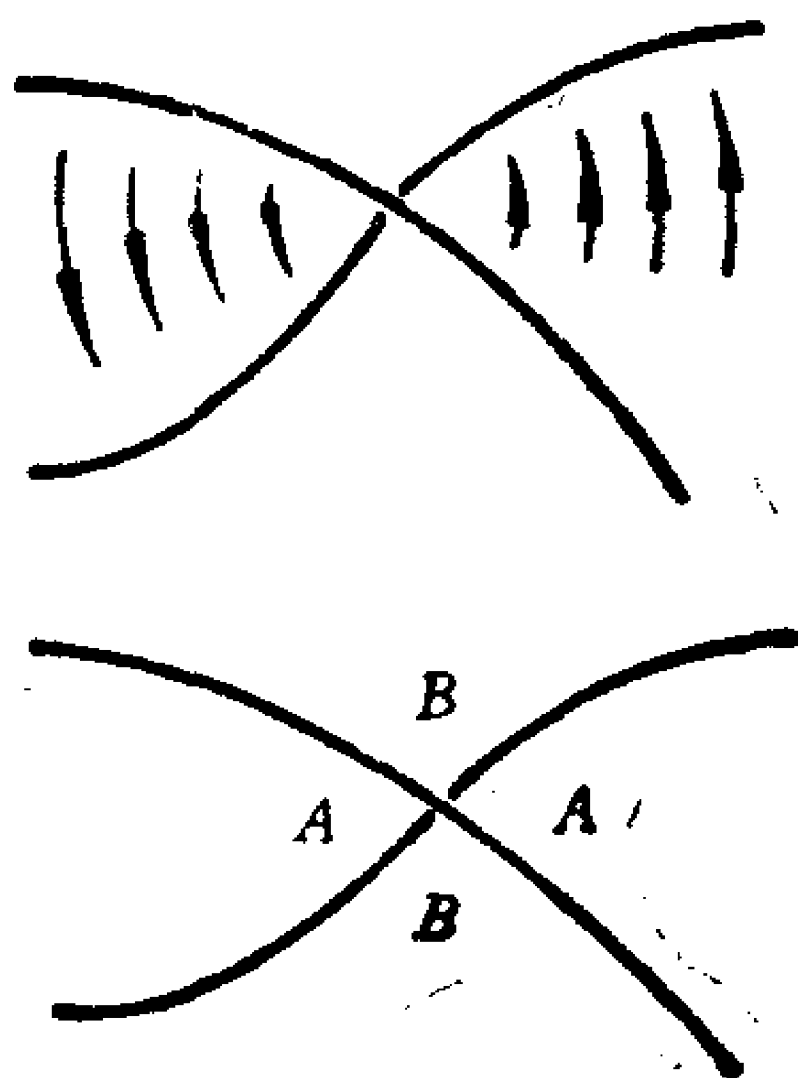
括号公理

$$(1)\ [\times]=A[\asymp]+B[\supset\subset],$$
$$[\times]=B[\asymp]+A[\supset\subset];$$
$$(2)\ [\bigcirc K]=d[K],$$
$$[\bigcirc]=d.$$

这两个公理的直观解释如下。

首先注意，一个非定向交叉在其顶点附近的四块区域分成两对。我们约定：把由上方曲线逆时针旋转至下方曲线所扫过的两个区域取成一对，标以记号 A；另一对区域标以 B（如下图所示）。于是，公理(1)中第一个公式就读成

$$[\times_{A,A}^{B,B}]=A[\asymp]+B[\supset\subset],$$



其中A对应于将A区域打通的切割，而B所对应的是将B区域打通的切割。在上述约定下，公理（1)中第二个公式的意义是明显的。这里应指出的一点是，这两个方程中的交叉，代表着包含它们的整个图形。也就是说 ⟩⟨，⟩⟨，⊃⊂ 要分别看成三个其他部分都相同的图形的组成部分。例如，公理(1) 中第一个展开式代表了诸如

$$[\ \cdot\ ]=A[\ \cdot\ ]+B[\ \cdot\ ]$$

的这些具体的公式。

公理(2) 是说，当图形中出现一个单独的圈时，其方括号多项式就等于去掉该圈后的图形的方括号多项式与d的乘积。特别地，[N条简单闭曲线]$=d^N$。例如

$$[\ \cdot\ ]=d^3.$$

显然，利用这两个公理通过逐步展开成简单闭曲线可以递归算出$[K]$的值。需要说明的是方括号多项式与（利用公理(1))打开交叉的次序无关，从而对每个图形来说是唯一确定的。为此，只要把$[K]$表示成在K的通用像U的所有状态上的和式。

设U是K的通用像，所谓U的一个状态就是在U的每个顶点选定一种切割的方式，在每个顶点的切割用记号表示出来(参见图14)。图 15 所表示的，是三叶结的通用像的一个状态及其所对应的切割。如果K有V个交叉，即其通用像U有V个顶点，则U的状态的数目就是 2^V 个。

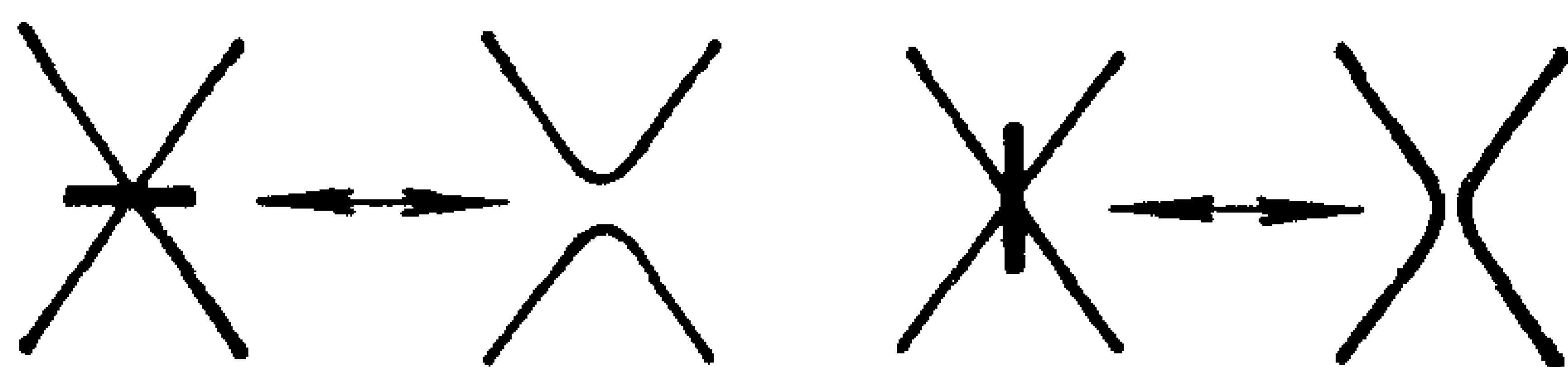


图14　顶点记号及其所对应的切割

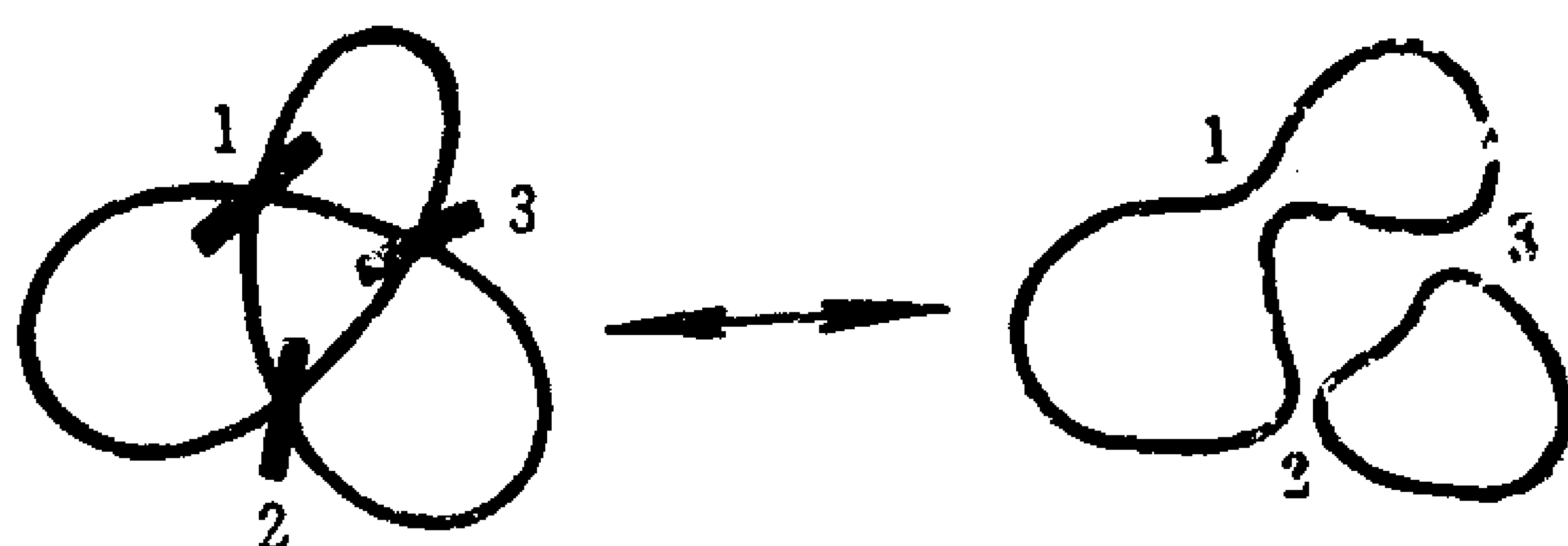
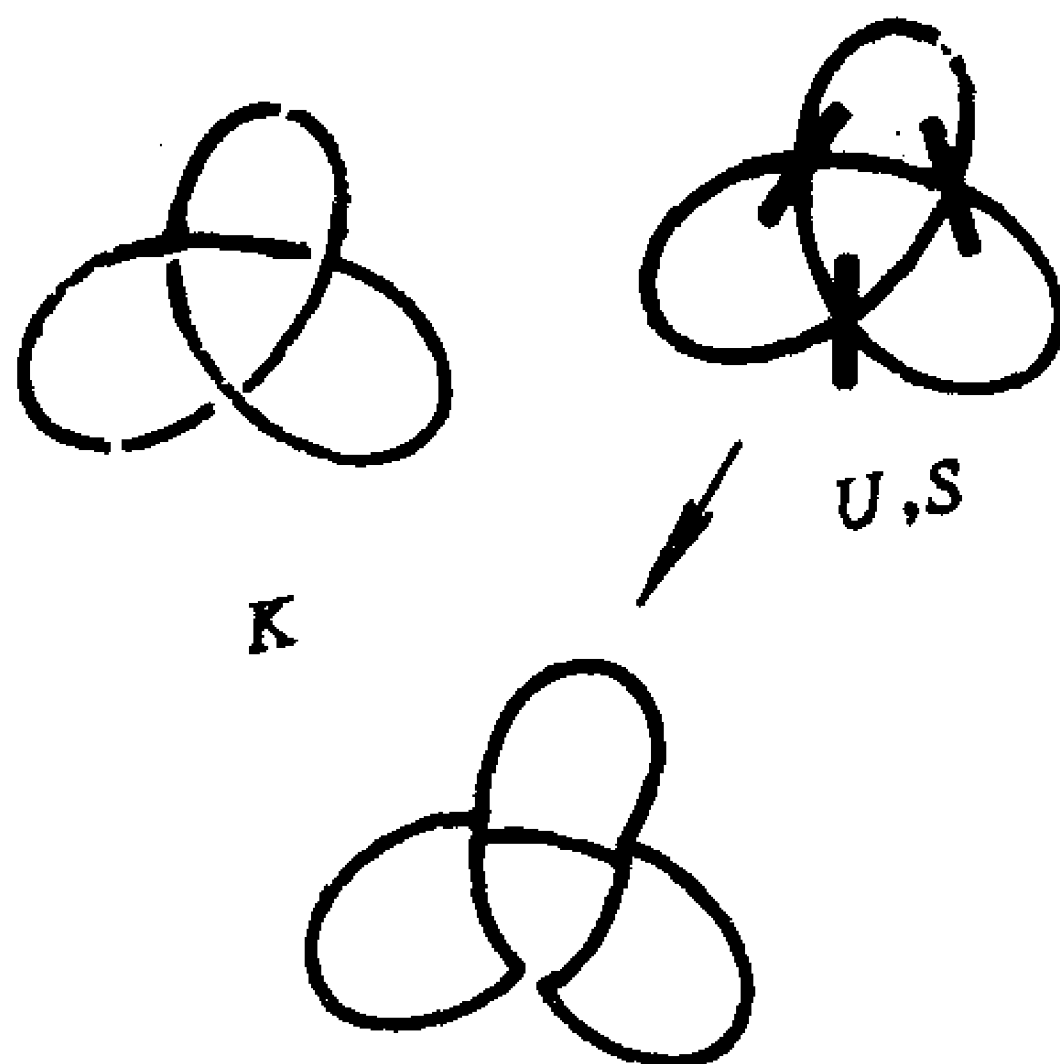


图15　状态及其所对应的切割方式

给定非定向图形K和它的通用像的一个状态S，我们总以$|S|$记通用像在S所对应的切割下所得到的连通分支（都是简单闭曲线）的数目，以$i_K(S)$记在S中打开A-通道的数目，以$j_K(S)$记在S中打开B-通道的数目。显然，K的交叉数目$V=i_K(S)+j_K(S)$，其中S为K的通用像的任一状态。例如在左图中有



$$i_K(S)=2,$$
$$j_K(S)=3-2=1;$$
$$|S|=1.$$

下述引理给出了图形的方括号多项式的唯一的值，它直接用括号公理以及三变元的可交换性展开而得到。它也可作为$[K]$的在逻辑上严密的定义。

引理2.1 $[K]=\sum_{S}A^{i_K(S)}B^{j_K(S)}d^{S}$，其中$\sum_{S}$表示在$K$的通用像的所有状态上求和。

2.2 不变多项式的衍生

建立方括号多项式以后，我们所关心的问题是，在什么条件下它可以衍生出拓扑不变量。它在平面合痕下显然是不变的。我们需要进一步考察它在R-运动下所产生的变化。

引理 2.2 $[\ \ ]=AB[\supset\subset]+(ABd+A^2+B^2)[\asymp]$。

证明
$$
\begin{aligned}
[\ \ ]&=A[\ \ ]+B[\ \ ]\\
&=A^2[\ \ ]+AB[\ \ ]\\
&\quad+BA[\supset\subset]+B^2[\asymp]\\
&=AB[\supset\subset]+(A^2+ABd+B^2)[\asymp].
\end{aligned}
$$

由此可见，如果$AB=1$，$d=-A^2-B^2$，则我们便得到[ ]在Ⅱ型R-运动下的不变性。

引理 2.3 若$[\ \ ]=[\supset\subset]$，则[ ]在Ⅲ型R-运动下也不变。

证明
$$
\begin{aligned}
[\ \ ]&=A[\ \ ]+B[\ \ ]\\
&=A[\ \ ]+B[\ \ ]\\
&=A[\ \ ]+B[\ \ ]\\
&=[\ \ ].
\end{aligned}
$$

本节以下我们总是假定 $B=A^{-1}$，$d=-A^2-A^{-2}$，并特别记

$$\langle K\rangle = d^{-1}[K],$$

于是相应的公理成为

(1) $\langle\ \ \rangle = A\langle\ \ \rangle + B\langle\ \ \rangle$,

$\langle\ \ \rangle = B\langle\ \ \rangle + A\langle\ \ \rangle$;

(2) $\langle \bigcirc K\rangle = d\langle K\rangle$,

$\langle \bigcirc \rangle = 1$.

这种特殊的括号$\langle\ \ \rangle$在 II，III 型 R-运动下是不变的。在 I 型 R-运动下，它的变化如下：

引理 2.4 令 $\alpha=-A^3$，则

$$\langle\ \ \rangle = \alpha\langle\ \ \rangle,$$

$$\langle\ \ \rangle = \alpha^{-1}\langle\ \ \rangle.$$

证明 $\langle\ \ \rangle = A\langle\ \ \rangle + A^{-1}\langle\ \ \rangle$

$= (A(-A^2-A^{-2})+A^{-1})\langle\ \ \rangle$

$= \alpha\langle\ \ \rangle$,

$\langle\ \ \rangle = A\langle\ \ \rangle + A^{-1}\langle\ \ \rangle$

$= (A+A^{-1}(-A^2-A^{-2}))\langle\ \ \rangle$

$= \alpha^{-1}\langle\ \ \rangle$.

我们称由 II，III 型 R-运动所生成的等价关系为正则合痕(regular isotopy)，则$\langle K\rangle$是 K 的正则合痕不变量。

在§1中已经知道，定向图形 K 的拧数 $w(K)$（即交叉符号之和）是一个正则合痕不变量。我们对定向的 K 定义

$$f_K = \alpha^{-w(K)}\langle K\rangle,$$

其中< >与定向无关（当K的分支数为1时，$w(K)$ 也与定向无关），则 f_K 在 I 型 R-运动下也是不变的，因而是K的一个外围合痕不变多项式，称之为Kauffman 多项式。

2.3 无旋向性的一个必要条件

注意到括号公理和拧数定义，下述结论是显然的。

引理2.5 对定向图形K以及反置其所有交叉而得到的镜像 K^*，多项式的取值有如下关系：

$$\langle K^*\rangle(A)=\langle K\rangle(A^{-1}),\quad f_{K^*}(A)=f_K(A^{-1}).$$

在这里，我们总是将$\langle K\rangle$看成A的一个表达式。如果用A^{-1}代替$\langle K\rangle$中的A，所得的表达式写成$\langle K\rangle(A^{-1})$；因此$\langle K\rangle$的原来的表达式也写成$\langle K\rangle(A)$。

推论 定向图形K无旋向的必要条件，是

$$f_K(A)=f_K(A^{-1}),$$

即 f_K 在形式上一定可以表达为

$$f_K=a_0+\sum_{\mu=1}^{m}a_\mu(A^\mu+A^{-\mu}),$$

其中$\{a_\mu|\mu=0,1,2,\cdots,m\}$是一组整数。

证明 K无旋向即$K\sim K^*$，从而由f_K的不变性得到$f_K=f_{K^*}$。故$f_K(A)=f_*(A)=f_K(A^{-1})$。证毕。

至此，我们可以证明三叶结的有旋向性，证法是迄今为止最为简便的。计算如下：

$$\langle \quad \rangle = A\langle \quad \rangle + A^{-1}\langle \quad \rangle$$
$$= A\alpha + A^{-1}\alpha^{-1}$$
$$= -A^4 - A^{-4},$$

$$\langle \quad \rangle = A\langle \quad \rangle + A^{-1}\langle \quad \rangle$$

T

$$= A(-A^4 - A^{-4}) + A^{-1}\alpha^{-2}$$
$$= -A^5 - A^3 + A^{-7},$$

$$w(T) = 3,$$

$$f_T = \alpha^{-3}\langle T\rangle = -A^{-9}(-A^5 - A^3 - A^{-7})$$
$$= A^{-4} + A^{-12} - A^{-16},$$

所以 $$f_T(A) = A^{-4} + A^{-12} - A^{-16}$$
$$\neq A^4 + A^{12} - A^{16} = f_T(A^{-1}),$$

从而T有旋向。

至于 Whitehead 环链W，同样可推导出$f_W(A)\neq f_W(A^{-1})$，即W是有旋向的，从而W是非平凡的环链。计算如下：

$w(W) = 0$，

$$\langle W\rangle = A\langle \quad \rangle + A^{-1}\langle \quad \rangle$$
$$= A^2\langle \quad \rangle + \langle \quad \rangle + A^{-1}\alpha^{-1}\langle \quad \rangle$$
$$= A^2 d + (1 - A^{-4})\Big(A\langle \quad \rangle + A^{-1}\langle \quad \rangle\Big)$$
$$= A^2(-A^2 - A^{-2}) + (1 - A^{-4})(A\alpha + A^{-1}\alpha^{-1})\langle \quad \rangle$$
$$= -A^4 - 1 + (1 - A^{-4})(-A^4 - A^{-4})^2$$
$$= A^8 - 2A^4 + 1 - 2A^{-4} + A^{-8} - A^{-12},$$

$$f_W = a^{-w(W)}\langle W\rangle$$
$$= A^8 - 2A^4 + 1 - 2A^{-4} + A^{-8} - A^{-12},$$
$$f_W(A) - f_W(A^{-1}) = -A^{-12} + A^{12},$$

其中W的图形如下图所示：



§3 交替纽结和环链

利用括号多项式，我们可以得到有关交替纽结和交替环链的一些精细的结果。其根据是，在确定出$\langle K\rangle$中A的最高次幂 max deg$\langle K\rangle$和最低次幂 min deg$\langle K\rangle$之后，这两者之差

$$\text{span}\langle K\rangle = \max\deg\langle K\rangle - \min\deg\langle K\rangle$$

是一个不变量。

3.1 Kauffman 幂次定理

一个环链（或纽结）称为交替的，如果它具有一个交替图形，也就是在每个分支上沿固定方向前进时，途经交叉处的方式是上-下-上-下…交替出现的（参见图16）。

如无特别声明，我们总是假设通用像是连通的。易见，如果把通用像所围成的区域黑白相间地染色（每个局部都像是国际象棋盘的一部分），那么对交替图形来说，在每个交叉处的一对黑色区域或者都是A-型的、或者都是B-型的（参见图16，这里的A,B与定义括号多项式时区域A,B的区分

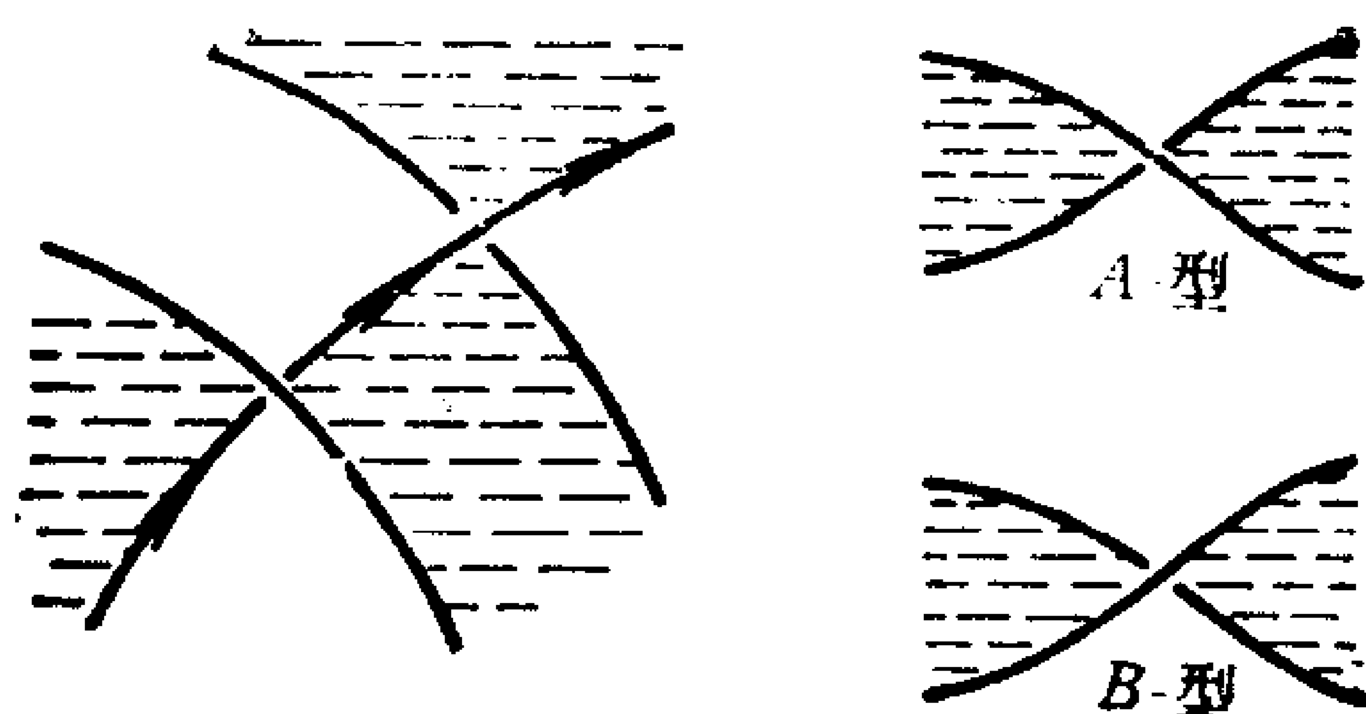


图16　交替图形

是一致的)。

我们知道，括号多项式是由如下的和式给出的：

$$\langle K\rangle=\sum_S A^{i_K(S)}A^{-j_K(S)}d^{|S|-1}$$

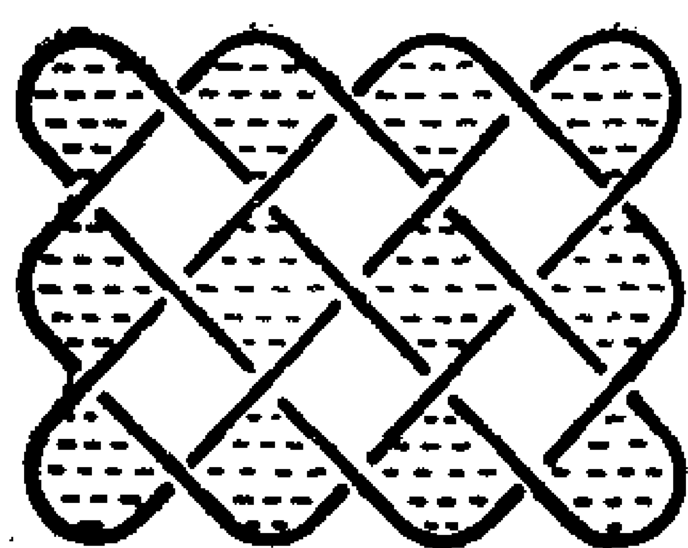
$$=\sum_S A^{i_K(S)-j_K(S)}(-A^2-A^{-2})^{|S|-1}.$$

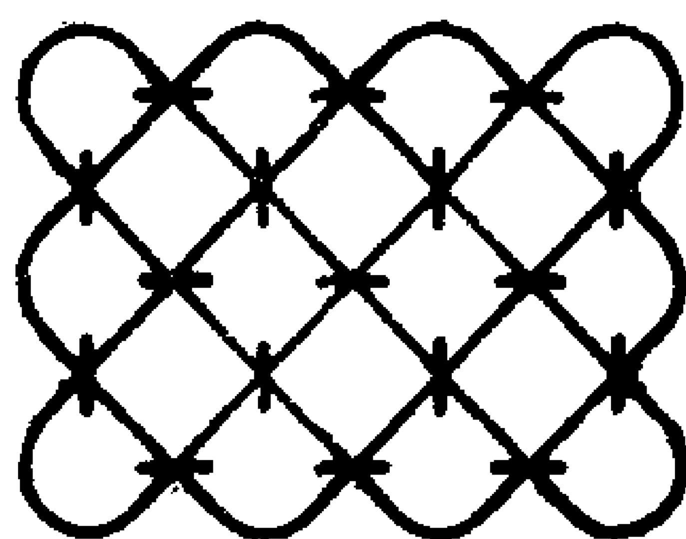
用 S_A表示所有的切割都是打开 A-通道的状态，S_B表示所有的切割都是打开 B-通道的状态，则

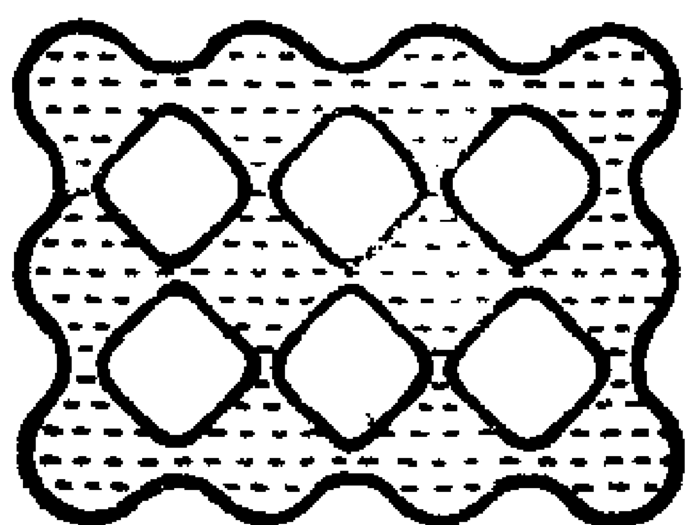
$$i_K(S_A)=V,\qquad j_K(S_A)=0,$$
$$i_K(S_B)=0,\qquad j_K(S_B)=V,$$

其中 V 是 K 的交叉数。对交替图形 K 而言，我们总可以采用 A-型染色（即 A-型区域染成黑色)，则在状态 S_A下所有的黑色区域打通成为一个大的黑色连通区域，并且该区域边界的每个分支恰好是相应的白色区域的边界（都是简单闭曲线)。记白色区域的数目是W，则$|S_A|=W$。参看图 17，其中 B 是黑色区域数目，R 是区域总数，V 是交叉数。于是，S_A所对应的项为 $A^V(-A^2-A^{-2})^{W-1}$。同理，S_B 所对应的项为

$$A^{-V}(-A^2-A^{-2})^{B-1}.$$







S_A-状态

$V=17$
$W=7, B=12$
$R=7+12$
$=V+2$
$|S_A|=7=W$

图17　状态 S_A

记 $\langle K|S\rangle = A^{i_K(S)-j_K(S)}d^{|S|-1}$，则 $\langle K\rangle = \sum\limits_{S}\langle K|S\rangle$．

$\langle K|S\rangle$的最高次项的幂次

$$\max\deg\langle K|S\rangle = i_K(S) - j_K(S) + 2(|S|-1),$$

最低次项的幂次

$$\min\deg\langle K|S\rangle = i_K(S) - j_K(S) - 2(|S|-1).$$

对交替图形 K，我们已经计算出

$$\max\deg\langle K|S_A\rangle = V + 2(W-1),$$

$$\min\deg\langle K|S_B\rangle = -V - 2(B-1).$$

为了估计其他状态的最高幂次和最低幂次，我们要建立下面的幂次比较定理。这里，通用像的任何一个状态 S，可以看成是从 S_A出发转换 $j_K(S)$ 个相应交叉点的切割方式形成的（同样，也可看成从 S_B出发转换 $i_K(S)$ 个相应交叉点的切

割方式形成的)。下述引理的证明在本节末给出。

引理3.1（幂次比较定理） 考虑环链（或组结）图形 K 及其通用像的一个状态 S。设 $i_K(S)\geqslant 1$.将 S 的一个 A-型切割换成 B-型切割，把所得到的状态记作 S'，则

(i) $i_K(S)-j_K(S)=i_K(S')-j_K(S')+2$;

(ii) $|S'|=|S|\pm 1$;

(iii)

$$\max\deg\langle K|S'\rangle=\begin{cases}\max\deg\langle K|S\rangle, & \text{当} |S'|=|S|+1;\\ \max\deg\langle K|S\rangle-4, & \text{当} |S'|=|S|-1.\end{cases}$$

由此，我们可归纳出如下的一般结论：

$$\max\deg\langle K|S_A\rangle=\max_S\{\max\deg\langle K|S\rangle\},$$
$$\min\deg\langle K|S_B\rangle=\min_S\{\min\deg\langle K|S\rangle\},$$

其中 S 为通用像的状态。为了能够确定 $\max\deg\langle K\rangle$ 和 $\min\deg\langle K\rangle$，我们讨论既约图形。

环链（或组结）的一个图形 K 称为既约图形，如果在 K 中不含有地峡（isthmus）。所谓地峡是指图形中的一个交叉，使得在它周围的四个局部区域中有两个实际上是整个图形的同一个区域的组成部分（如图 18 所示）。下面的幂次比较定

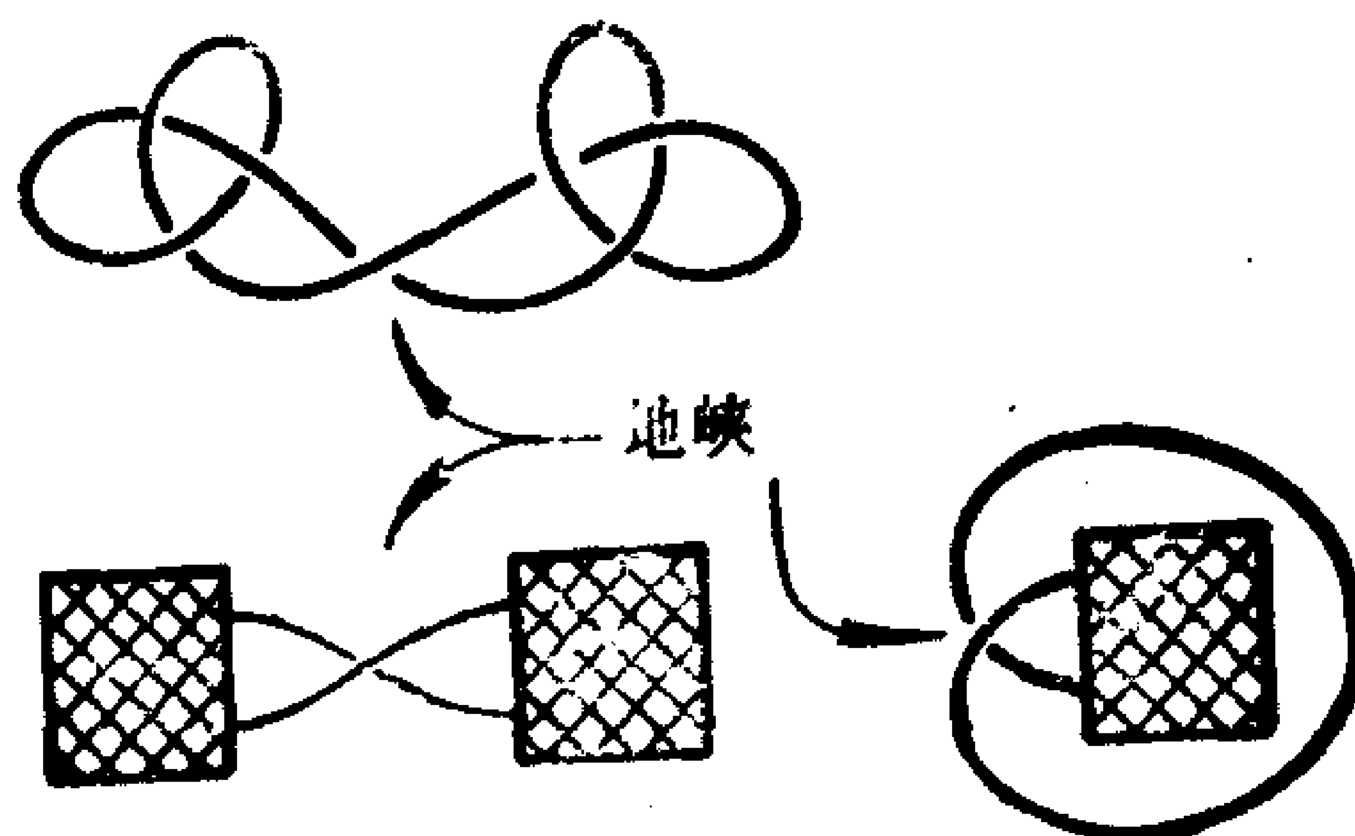


图18 地峡

理是既约图形所特有的，其证明也放在后面叙述。

引理3.2 令 K 是连通的既约图形，S 是由 S_A 转换一个切割方式所形成的状态，则 $|S| = |S_A| - 1$，从而

$$\max\deg\langle K|S\rangle = \max\deg\langle K|S_A\rangle - 4.$$

至此，我们可以证明 Kauffman 幂次定理。

定理3.1 (Kauffman) 令 K 为连通既约交替图形，则

$$\max\deg\langle K\rangle = V + 2(W-1),$$

$$\min\deg\langle K\rangle = -V - 2(B-1),$$

其中 V 是交叉的数目，W 和 B 分别是在 A-型染色下的白色和黑色区域的数目，并且在 $\langle K\rangle$ 中最高次项、最低次项的系数是 ± 1。

证明 由引理 3.1 和引理 3.2知道，$\langle K\rangle$ 中 A 的最高次项就是 $\langle K|S_A\rangle$ 中 A 的最高次项 $A^V(-A^2)^{W-1} = (-1)^{W-1} \times A^{V+2(W-1)}$，从而有关最高次项的结论成立。关于最低次项可通过建立与引理 3.2 平行的结论得到证明。证毕。

注 关于定理 3.1 中最低次项的证明也可以利用对偶性考察镜像 K^* 的最高次项而获得。

3.2 图形的交叉数

下述定理是幂次定理的重要推论。

定理3.2 (Kauffman-Murasugi-Thistlethwaite) 交替环链（或纽结）L 的既约交替图形的交叉数，是 L 的拓扑不变量。

证明 首先，Kauffman 多项式是定向图形的外围合痕不变量，从而 $\langle L\rangle$ 的最高次幂与最低次幂的差是与定向无关的拓扑不变量，记之为 span(L)。

设 L_0是 L 的既约交替图形，则显然有

$$\begin{aligned}\operatorname{span}(L) &= \operatorname{maxdeg}\langle L_0\rangle - \operatorname{mindeg}\langle L_0\rangle \\ &= (V+2(W-1))-(-V-2(B-1)) \\ &= 2V+2(W+B-2) \\ &= 2V+2(R-2),\end{aligned}$$

其中 V,W,B,R 分别是 L_0的交叉数，白色区域数，黑色区域数和区域总数。注意到4-价平面图的边数是顶点数的二倍，由连通平面图的 Euler 公式即知道 $R=V+2$。因此，

$$\operatorname{span}(L)=4V,$$

从而 $V=\frac{1}{4}\operatorname{span}(L)$ 是 L 的拓扑不变量。证毕。

关于既约交替图形的存在性，我们要作些说明。由地峡的定义知道，在它周围的一对局部区域是整个图形中一整块区域（记之为D）的组成部分（见图19）；显然，存在相应的

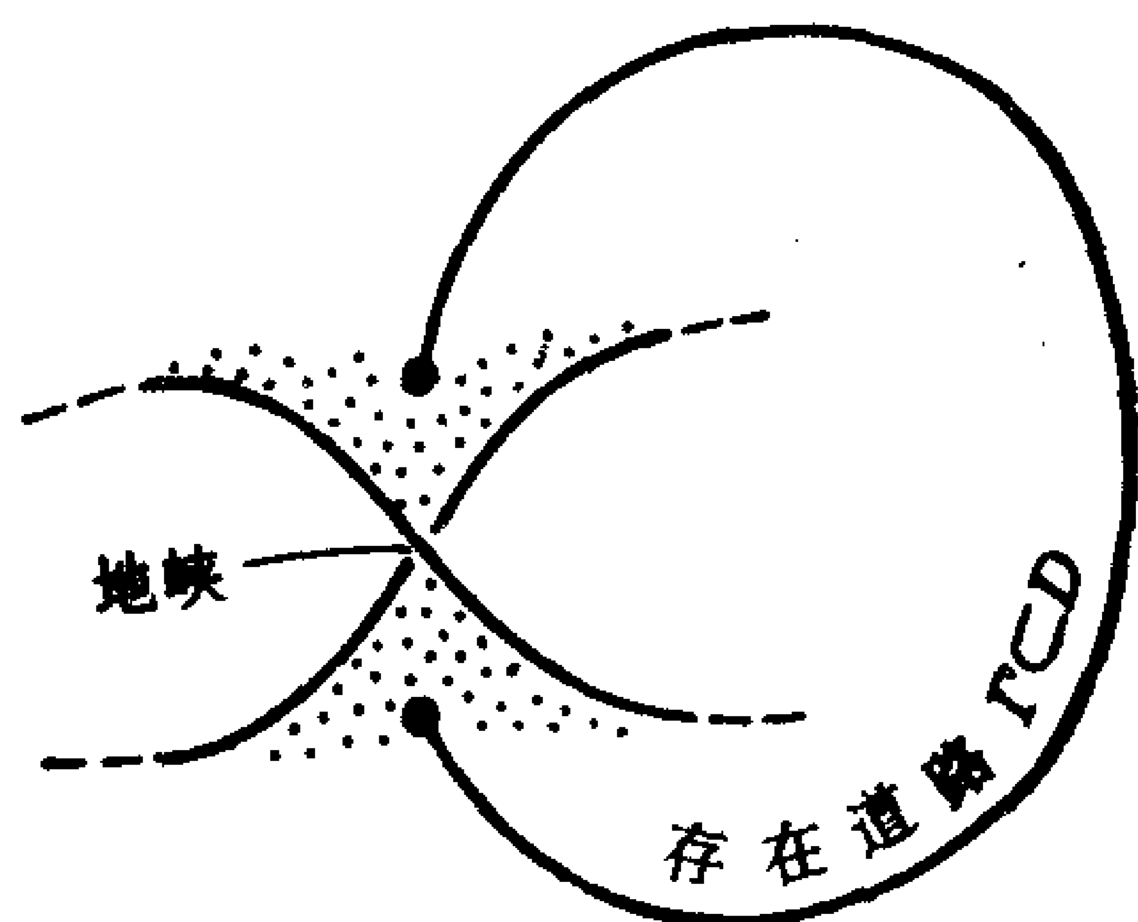


图 19

外围合痕使这个地峡消失（事实上只要把图形相应的部分按适当的方向整个翻转180°即可）。所有的图形 K 都可以按上述

方式外围合痕等价于相应的既约图形；并且，如果K是交替图形，则得到的是既约交替图形。因此，一个交替环链（或纽结）的既约交替图形总是存在的。此外应注意到在上述约化过程中不会增加交叉的个数。

定理 3.2 所解决的，是 Tait 和 Little 于上世纪末提出的经典的猜想之一。他们同时还猜测：交替环链（或纽结）K的既约交替图形K_0的交叉数不会超过K的任何一个图形的交叉数。下面我们来解决这个猜想。

引理3.3 对连通的通用像U及其任一状态S，记$\hat{S}$为将S的各个切割方式全部反置所得到的状态。则有

$$|S|+|\hat{S}|\leqslant R,$$

其中R是U的区域数。

该引理的证明在本节最后叙述。图 20 给出了具体的例子。

定理3.3 对任何既约图形K，有$\mathrm{span}(K)\leqslant 4V$，其中$V$是$K$的交叉数。

证明 由引理 3.1 已经知道

$$\max\deg\langle K\rangle\leqslant V+2(|S_A|-1),$$

$$\min\deg\langle K\rangle\geqslant -V-2(|S_B|-1).$$

而引理3.3说明$|S_A|+|S_B|\leqslant R=V+2$，故

$$\begin{aligned}\mathrm{span}(K)&=\max\deg\langle K\rangle-\min\deg\langle K\rangle\\&\leqslant 2V+2(|S_A|+|S_B|-2)\\&\leqslant 4V.\end{aligned}$$

证毕。

显然，由定理3.2的证明过程以及定理3.3可见，在一个交替环链(或纽结)L的所有图形中，既约交替图形的交叉数

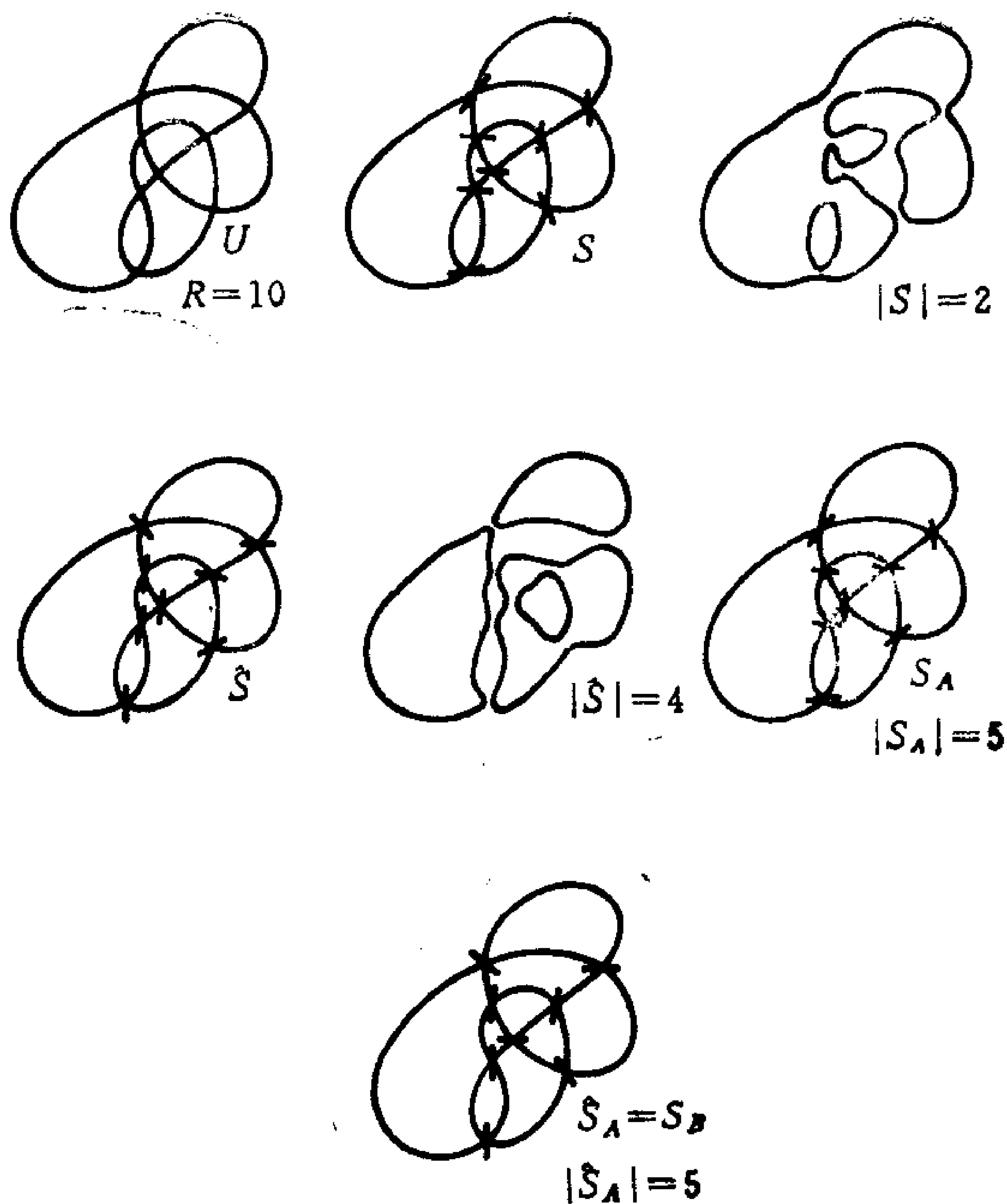


图20 $|S|+|\hat{S}|\leqslant R$

达到最小，并且是L的拓扑不变量。

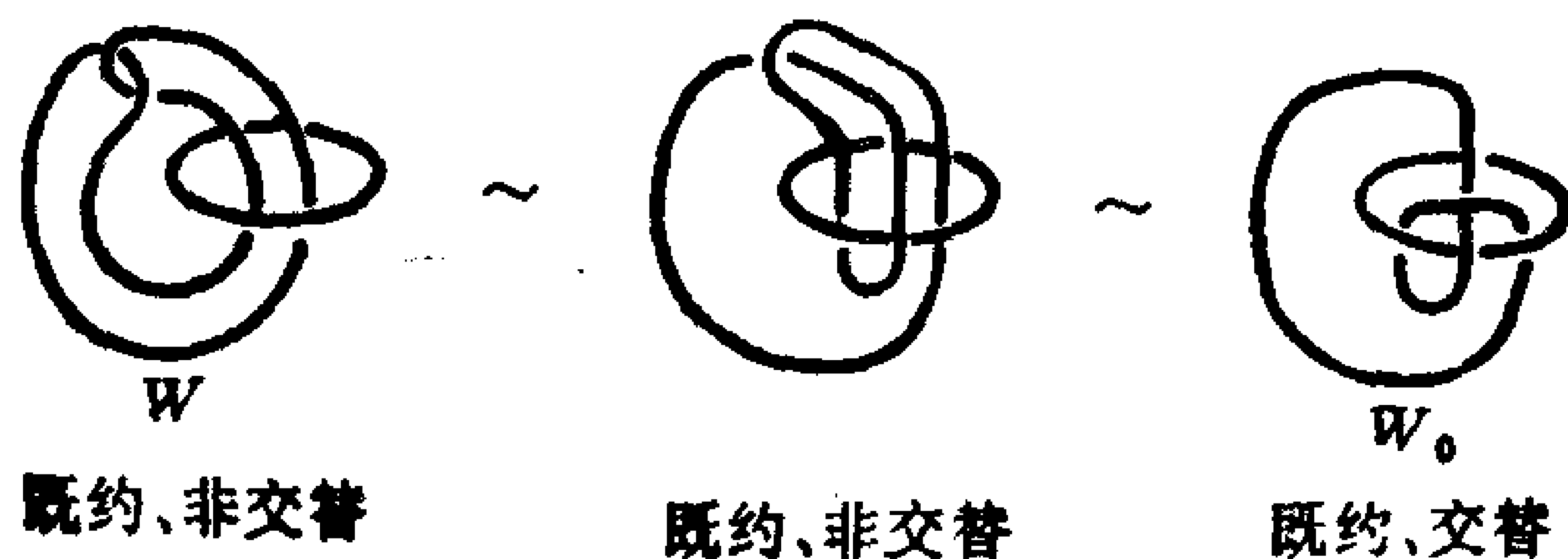
下面观察两个实例。

例 1



$$\operatorname{span}(L_1)=\operatorname{span}(L_2)=4\times4=16.$$

例 2 对 Whitehead 环链W，有 $\operatorname{span}=8-(-12)=20=4\times5$。事实上，$W$的既约交替图形 W_0具有五个交叉：



既约、非交替　　既约、非交替　　既约、交替

可见，非交替的既约图形的交叉数不一定是拓扑不变量。对一般的连通既约图形K，吴英青证明了如下结论："$\operatorname{span}(K)=4V$"当且仅当$K$是所谓有限个既约交替图形的"和"（参见 Y.Q.Wu, Jones polynomial and the crossing number of links, Lecture Notes in Math., vol. 1369, p.286—288 (1989)）。

关于既约交替图形，Tait 还提出如下的"翻转"猜想：同一交替环链（在外围合痕意义下）的两个既约交替图形可以由翻转而互相得到。这里所谓的翻转是指图形中有两个输入端和两个输出端的组结块作 180°的转动（如图21 所示）。翻转猜想蕴含着既约交替图形K的拧数$w(K)$是一个外围合痕不变量。后者已经由 Morwen Thistlethwaite 作出证明，但是翻转猜想本身仍是一个未解决的问题。

3.3 镜像

幂次定理可用来考察交替环链（或组结）的旋向性。

从引理 2.5 推论知道，对无旋向图形K的任何一种定向

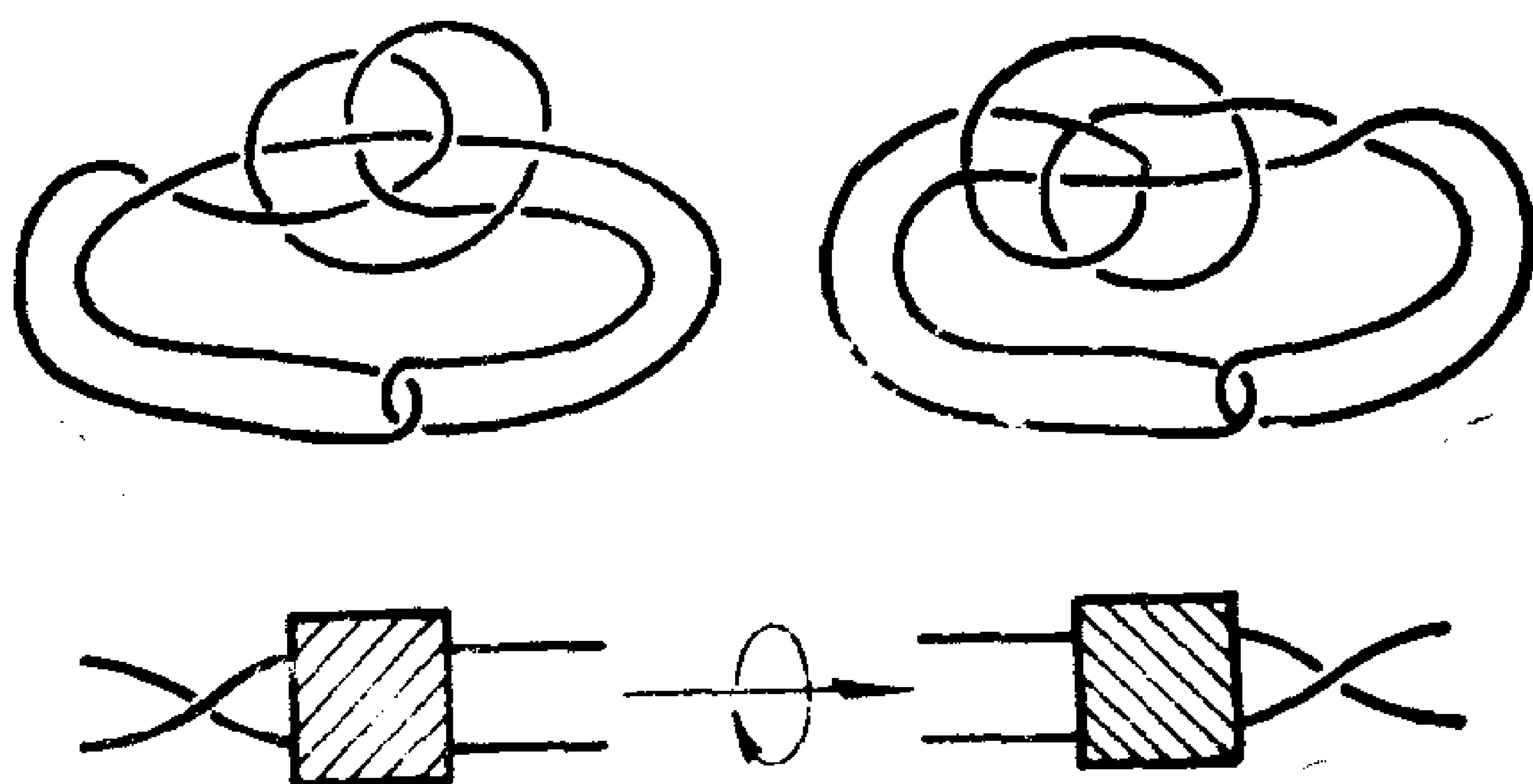


图21 翻转

总成立

$$\max \deg f_K = -\min \deg f_K.$$

于是 Kauffman 幂次定理告诉我们，对连通既约交替无旋向图形K定向后，总有 $-3w(K)+V+2(W-1)=-[-3w(K)-V-2(B-1)]$，即

$$3w(K)=W-B.$$

由此可见，$w(K)$ 与定向无关，从而f_K也与定向无关。利用 $W+B=V+2$，上述必要条件可化为

$$w(K)=\frac{V}{3}-\frac{2}{3}(B-1)=-\frac{V}{3}+\frac{2}{3}(W-1).$$

所以，当连通的既约交替图形K满足$|w(K)|\geqslant\frac{V}{3}$时，$K$是有旋向的。

事实上，Tait 的下述猜想是正确的：

定理3.4 (Thistlethwaite) 设K是一个既约交替图形。若K无旋向，则拧数 $w(K)=0$。

无旋向性是拓扑不变性质。定理 3.4 说明，既约交替无

旋向图形的交叉数一定是偶数．从这个观点来看，Whitehead环链的有旋向性是一目了然的．

既约交替无旋向图形是很多的．Kauffman猜测：每个这样的图形K不仅满足$W=B=\frac{V}{2}+1$，而且白色区域的伴随平面图G_W同构于黑色区域的伴随平面图G_B（如图22所示）．

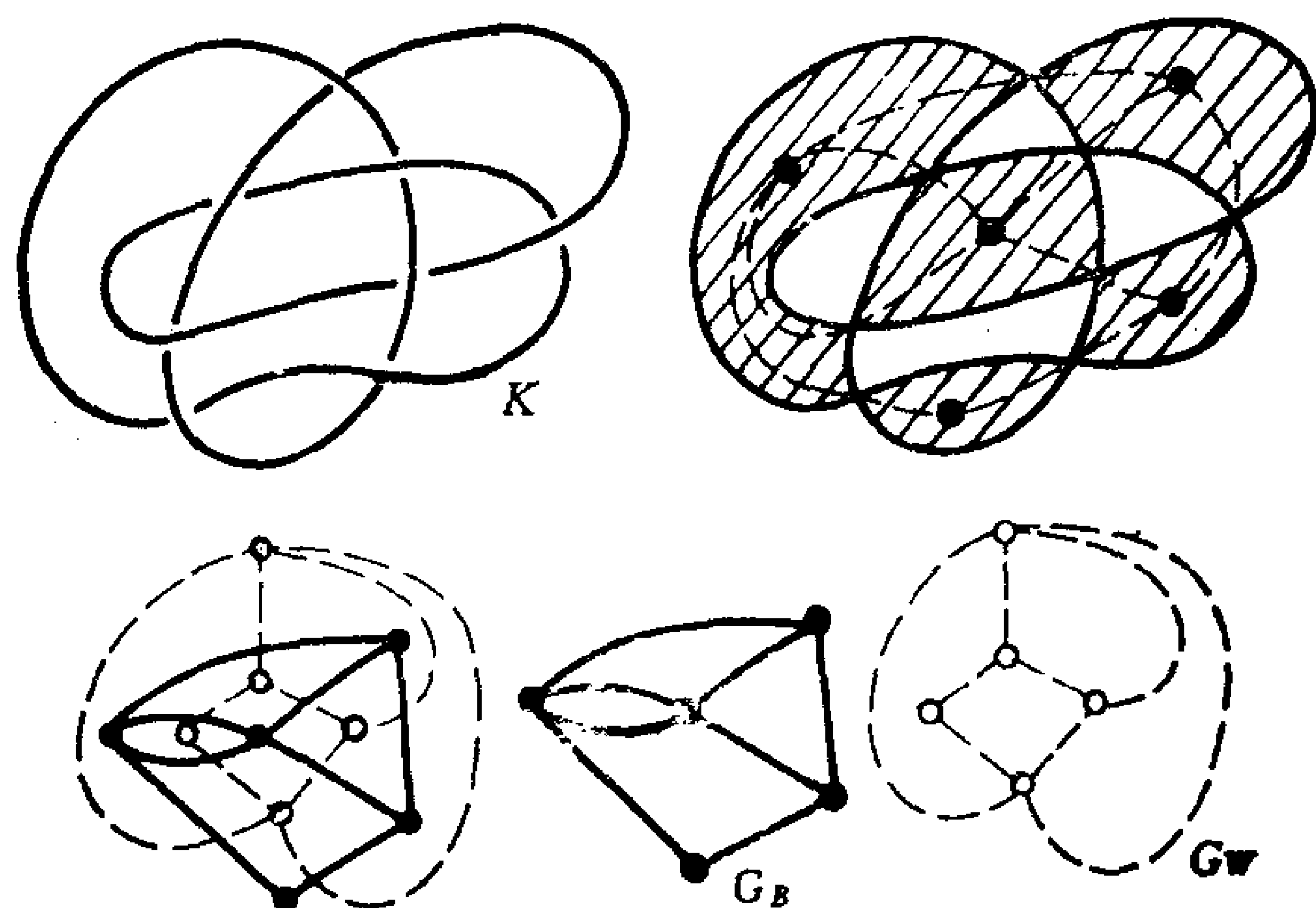


图22 伴随图同构

3.4 引理的证明

引理3.1的证明 结论(i)显然．结论(iii)由(i)，(ii)即得．所以只需要证结论(ii)．

我们知道，$|S|$是在S所对应的切割之后所形成的简单闭曲线的数目．现在S'和S的唯一不同之处是在某一个交叉处的切割方式不同（见图23）．观察那个特殊的交叉．如果以S的方式切割后，在该交叉处切割成的两条曲线段位于两

条不同的闭曲线上，则以 S' 的方式切割的结果就可看成将这两条闭曲线合二为一，而保持其它各条闭曲线不变，即：$|S'|=|S|-1$。反过来，如果以 S 的方式切割后，在该交叉处形成的两条（局部）曲线段实际上是同一条闭曲线的组成部分，则以 S' 的方式切割的结果就是将此闭曲线一分为二而保持其它各条闭曲线不变，即有 $|S'|=|S|+1$。故结论（ii）成立，引理证毕。

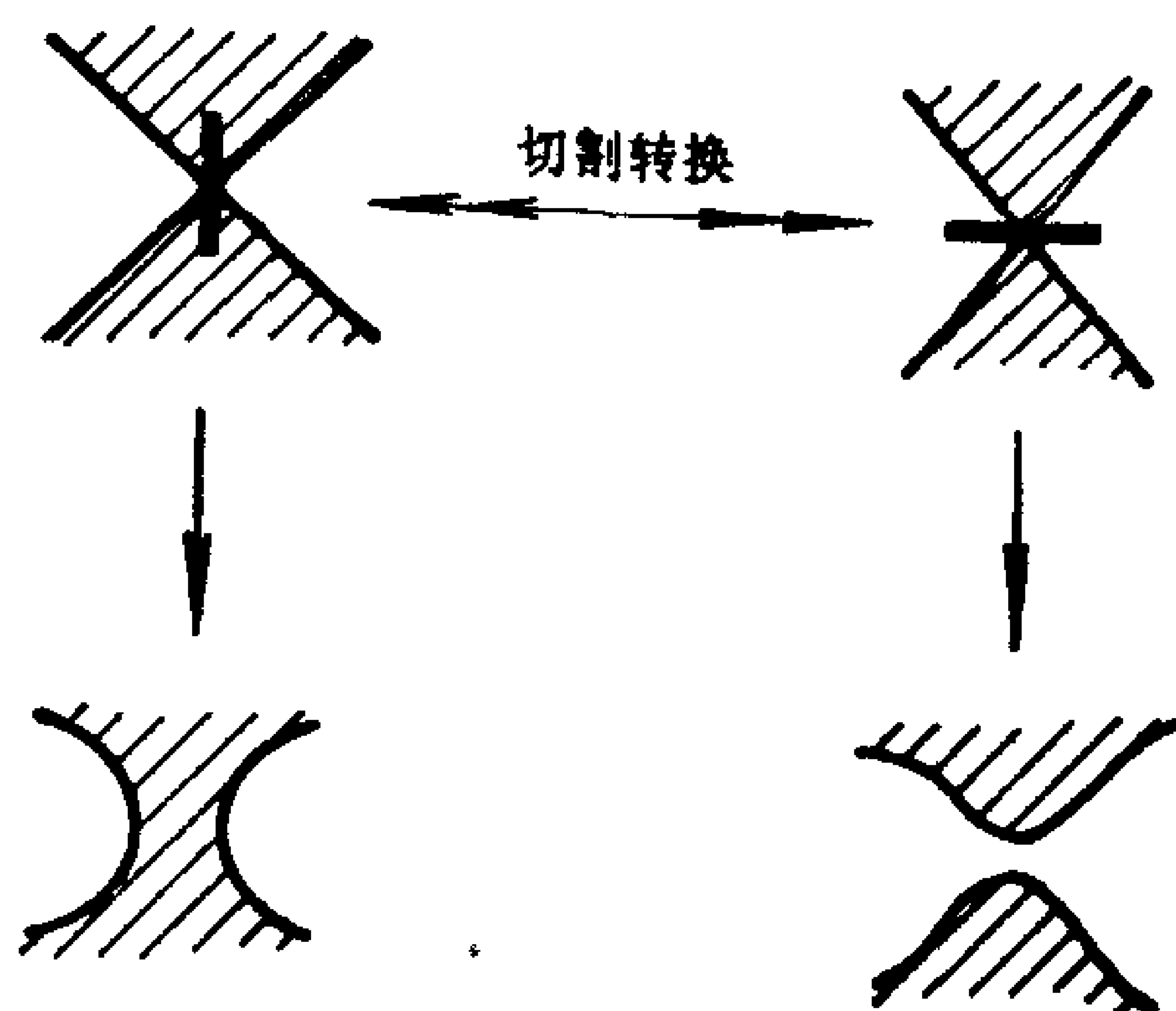


图 23

引理3.2的证明 用反证法。由引理 3.1，当 $|S|\neq|S_A|-1$ 时，只能有 $|S|=|S_A|+1$。此时，转换切割方式的那个交叉处在 S 的切割下所成的两条（局部）曲线段，分别落在两条不同的简单闭曲线上。这两条闭曲线的可能的位置关系有两种：(i) 一条落在另一条的内部区域内；(ii) 互在对方的外部区域内。如图24所示，黑色区域是在状态 S_A 下所打通的区域，那么在状态 S 下所形成的两条闭曲线间的白色区域 D 也只有两种可能性：(i) D 夹在两条闭曲线之间；(ii)

D落在两条闭曲线外部区域的公共部分之中。由 S_A的构造可见，在状态 S_A的切割方式下在该交叉处所形成的两个局部白色区域是一整块白色区域的组成部分。即那个交叉是一个地峡，这与K是既约图形相矛盾。所以只能有 $|S| = |S_A| - 1$。由引理3.1的(iii)便有

$$\max \deg\langle K|S\rangle = \max \deg\langle K|S_A\rangle - 4.$$

证毕。

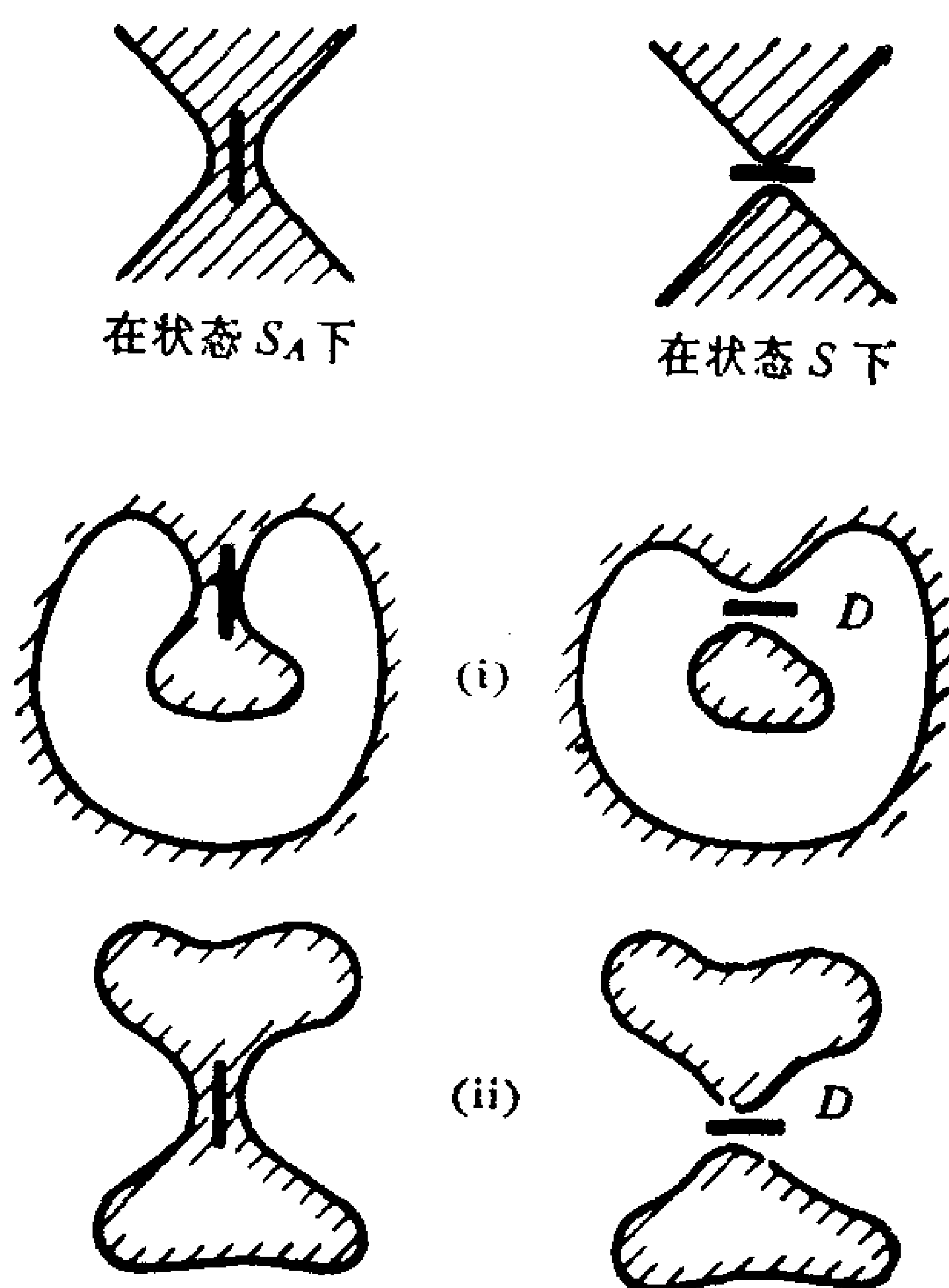
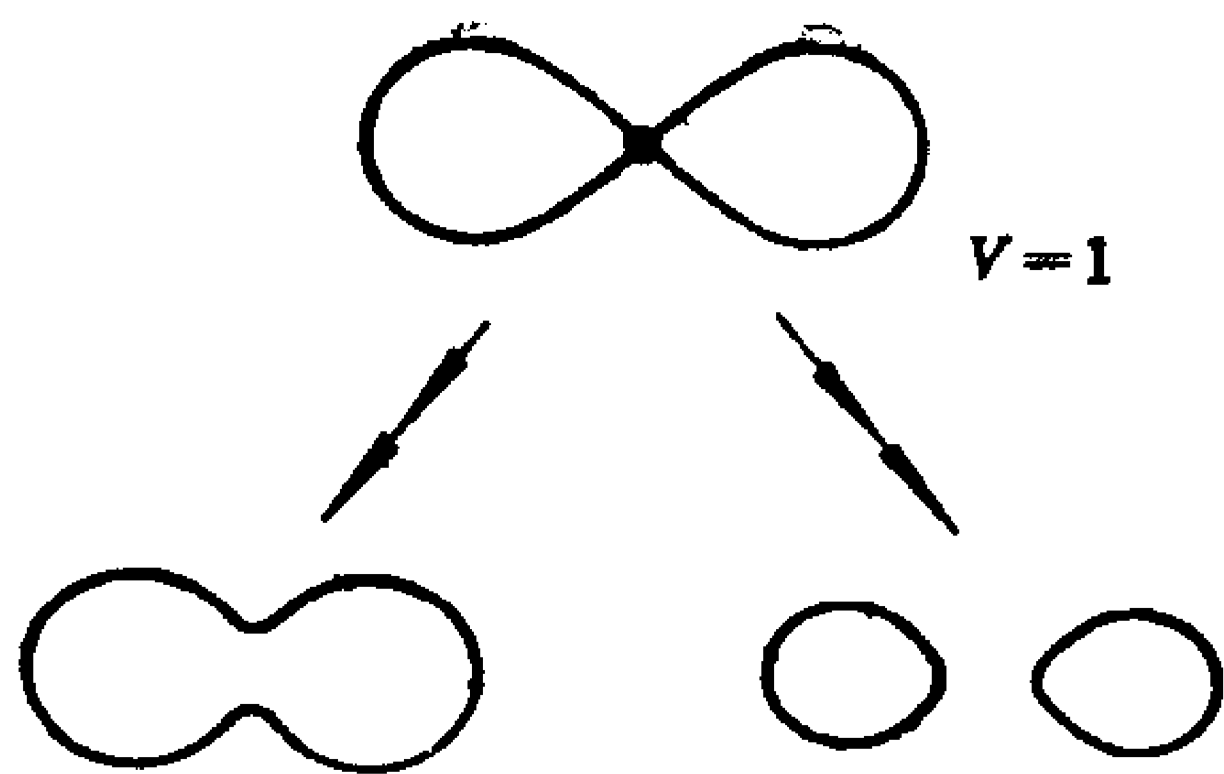


图 24

引理3.3的证明 对U的顶点数 $V = R - 2$ 用归纳法。

若 $V = 1$，则

$$|S| + |\hat{S}| = 1 + 2 = 3 = R \leqslant R.$$



当$V\geqslant 2$时，设结论对具有$(V-1)$个顶点的连通通用像成立，要证明相应的不等式对具有V个顶点的连通通用像U也成立。取定U的一个顶点v_0，对它进行切割，则得到两个相应的具有$(V-1)$个顶点的通用像U_1和U_2（见图25），其中至少有一个是连通的。不妨设U_1连通，则由归纳假定，成立

$$|S_1|+|\hat{S}_1|\leqslant R_1=V_1+2=(V-1)+2=R-1,$$

其中S_1是S在U_1上的限制；而$\hat{S}_1$作为S_1的反置，也是$\hat{S}$在

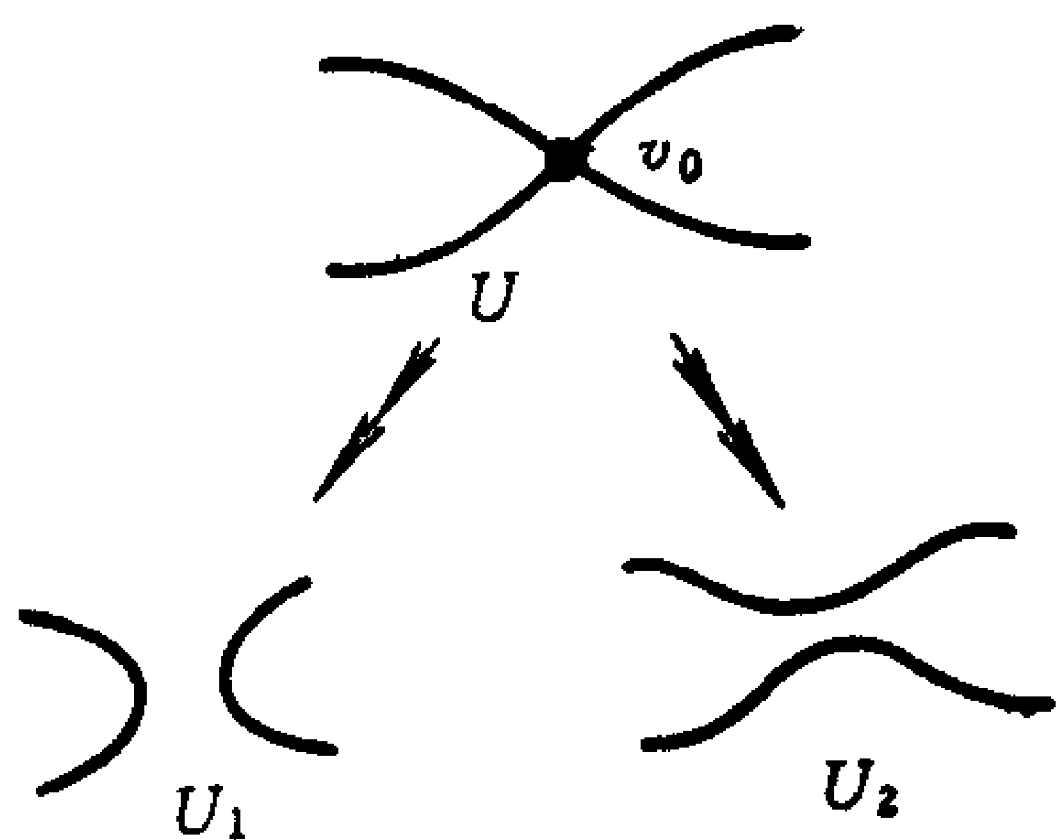


图 25

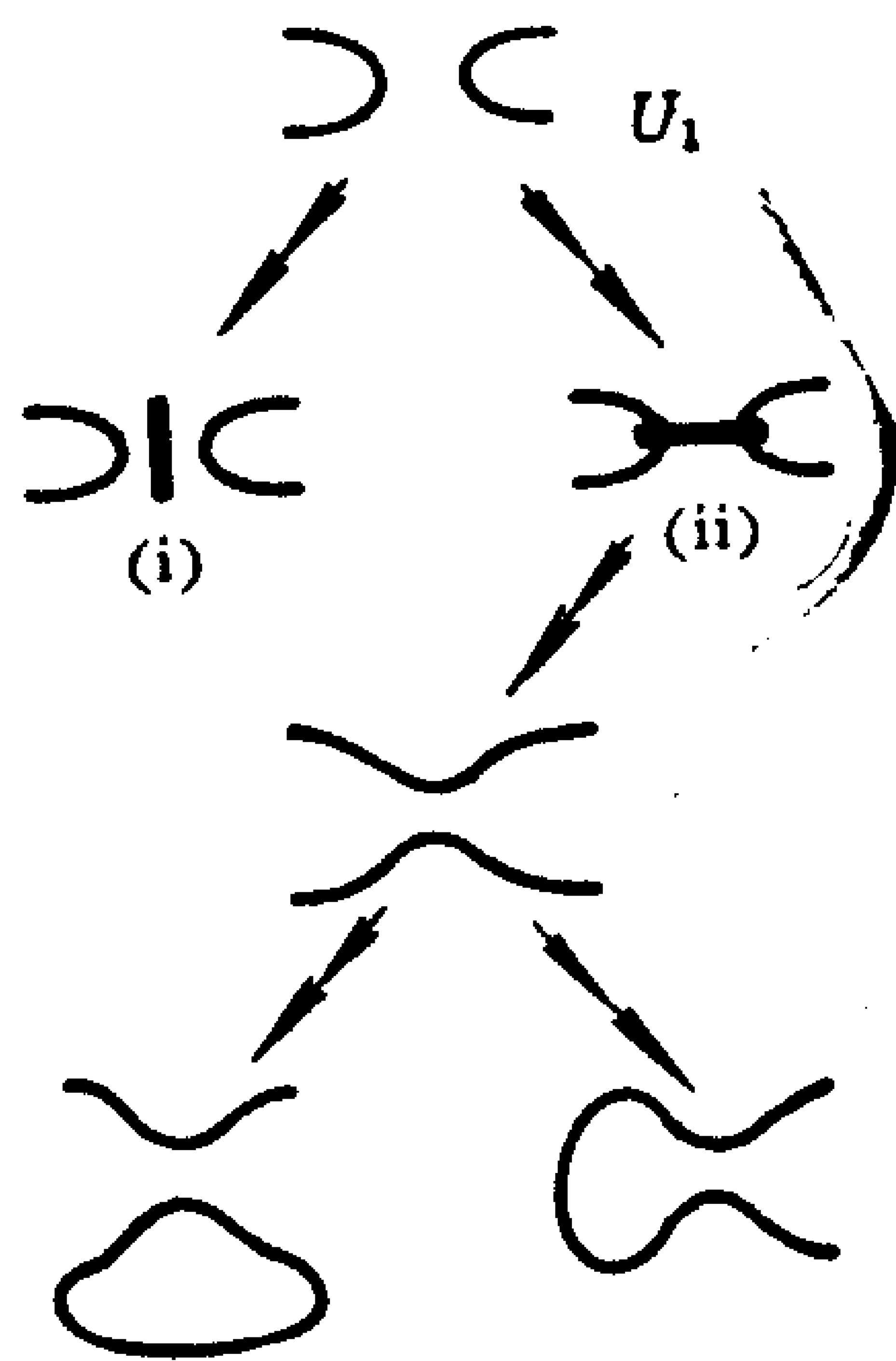


图 26

U_1 上的限制。现在反过来看，S 所对应的切割结果（在 U 上进行），就是在 S_1 所对应的切割结果（在 U_1 上进行）的基础上进一步加工而得到的；具体说即可分两种情形：

(i) 不再变动；

(ii) 或者将 S_1 下的两条闭曲线合二为一，或者将 S_1 下的一条闭曲线一分为二（参见图 26）。在情形 (i)下，$|S|=|S_1|$；在情形 (ii)下，$|S|=|S_1|\pm 1$。同理可分析 $\hat{S}$，并注意与 S 的相应情形进行对照，得知：在 $|S|=|S_1|$ 时 $|\hat{S}|=|\hat{S}_1|\pm 1$，而 $|S|=|S_1|\pm 1$ 时 $|\hat{S}|=|\hat{S}_1|$。于是无论何时总有

$$|S|+|\hat{S}|\leqslant|S_1|+|\hat{S}_1|+1\leqslant(R-1)+1=R.$$

这样，由归纳法完成证明。

（陈维桓 审校）

用2维图像法解高维线性规划问题[①]

W.P.Cooke

本文用一些具体例子说明，如何用2维图像法去解含有3个以上变量的线性规划问题。所用的方法很独特，当然，并不能用它去解所有的线性规划问题。但是，作为一种教学手段，它是有价值的。而且可作为一种“猜想和创见”，去推动大学生的科学研究。而且，这里用到的数学知识并不深，高中班的数学兴趣小组完全可以接受。

读者将会看到，这一图解法并不能取代单纯形法，后者是解一般的线性规划问题的一种很有效的计算方法。它的主要价值在于：对于几乎每一个只要会画直线的人，它都会触动他（她）的思想。

1. 技巧的理论依据

我们的做法基于以下2点基本的考虑。首先，空间中的任意一点，如果满足2个不等式，则必满足它们的线性组合，当然，要求所得到的不等式的意义是明了的。其次，画出一个n元方程的图形，并不一定要用n维空间。

---

① Two-dimensional graphical solution of higher-dimensional linear programming problem, *Math.Magazine*, 1973, 3—4, 70—76.

2．3个变量的例子

考虑下述线性规划问题。

例 1　求 $Z=2x_1+2x_2+x_3$ 的最大值，约束条件是：

$$
\begin{aligned}
&(a)\ x_1+x_2+x_3\leqslant 12,\\
&(b)\ x_1+2x_2-x_3\leqslant 5, \qquad x_1\geqslant 0, x_2\geqslant 0, x_3\geqslant 0. \qquad (1)\\
&(c)\ x_1-x_2+x_3\leqslant 2,
\end{aligned}
$$

这是教科书上的一个典型例子，它的可行解显然存在，但最优解并不明显。为求最优解，需作3次常规的单纯形迭代，或者用3维空间的图解法。

我们用2维空间的图解法来解。令 $X=2x_2+x_3$。代入后，目标函数可简单地写成 $Z=2x_1+X$。但约束条件的情形却不是如此简单。

接着，将约束条件（b）的 k 倍加到（a）上去，其中的 k 要选取得使所得不等式中 x_2 与 x_3 的系数之比等于2:1。因为涉及的是不等式而不是方程，所以仅仅需要仔细的是，k 要使得组合而得的不等式的意义是能确定的。这样做了之后，变量 X 就能明确地进入约束条件组。

勇敢地毫不顾忌到产生的可行解区域可能与（1）中原始的可行解区域不相等，我们将（a），(b)和(c）进行组合，得到只含变量 X 和 x_1 的两个不等式。先选取 k 使得

$$\frac{1+2k}{1-k}=\frac{2}{1} \text{ 或者 } k=\frac{1}{4}, \qquad (2)$$

组合(a)和(b)得

$$(x_1+x_2+x_3-12)+\frac{1}{4}(x_1+2x_2-x_3-5)\leqslant 0, \qquad (3)$$

整理得

$$5x_1+3X\leqslant 53. \qquad (4)$$

注意，(1) 的任意一个可行解，因为满足约束条件 (a) 和 (b)，所以必满足不等式 (4)。

类似地，选取 k 使得

$$\frac{2-k}{-1+k}=\frac{2}{1} \text{ 或者 } k=\frac{4}{3}, \qquad (5)$$

由 $k\cdot(c)+(b)$ 得

$$7x_1+X\leqslant 23. \qquad (6)$$

可以看到，若 (2) 和 (5) 中的某一个 k 是负数，则相应所得的不等式的意义将会是不清楚的。如果出现这种情形，就意味着，或者作某一个不同的替换（例如 $Y=x_1+x_2$）也许有用，或者（更糟糕的）是作任何替换都没有用。也就是说，并不是总能把原始问题简化成只含两个变量的问题。

成功地得到了 (4) 和 (6) 后，现在考虑下面的简化问题：

例 1′ 求 $Z_1=2x_1+X$ 的最大值，约束条件是

$$5x_1+3X\leqslant 53, \quad 7x_1+X\leqslant 23, \qquad (7)$$

$$x_1\geqslant 0, \quad X\geqslant 0.$$

由它的构造方法知，(1) 的任何一个可行解都是 (7) 的可行解。因此，如果 (7) 有一个有限的最优解，设为 Z_1^*，则必有 $Z_{max}\leqslant Z_1^*$。进而，明显的事实是：如果 (1) 有一个可行解 $Z=Z_1^*$，则它必是 (1) 的最优解。

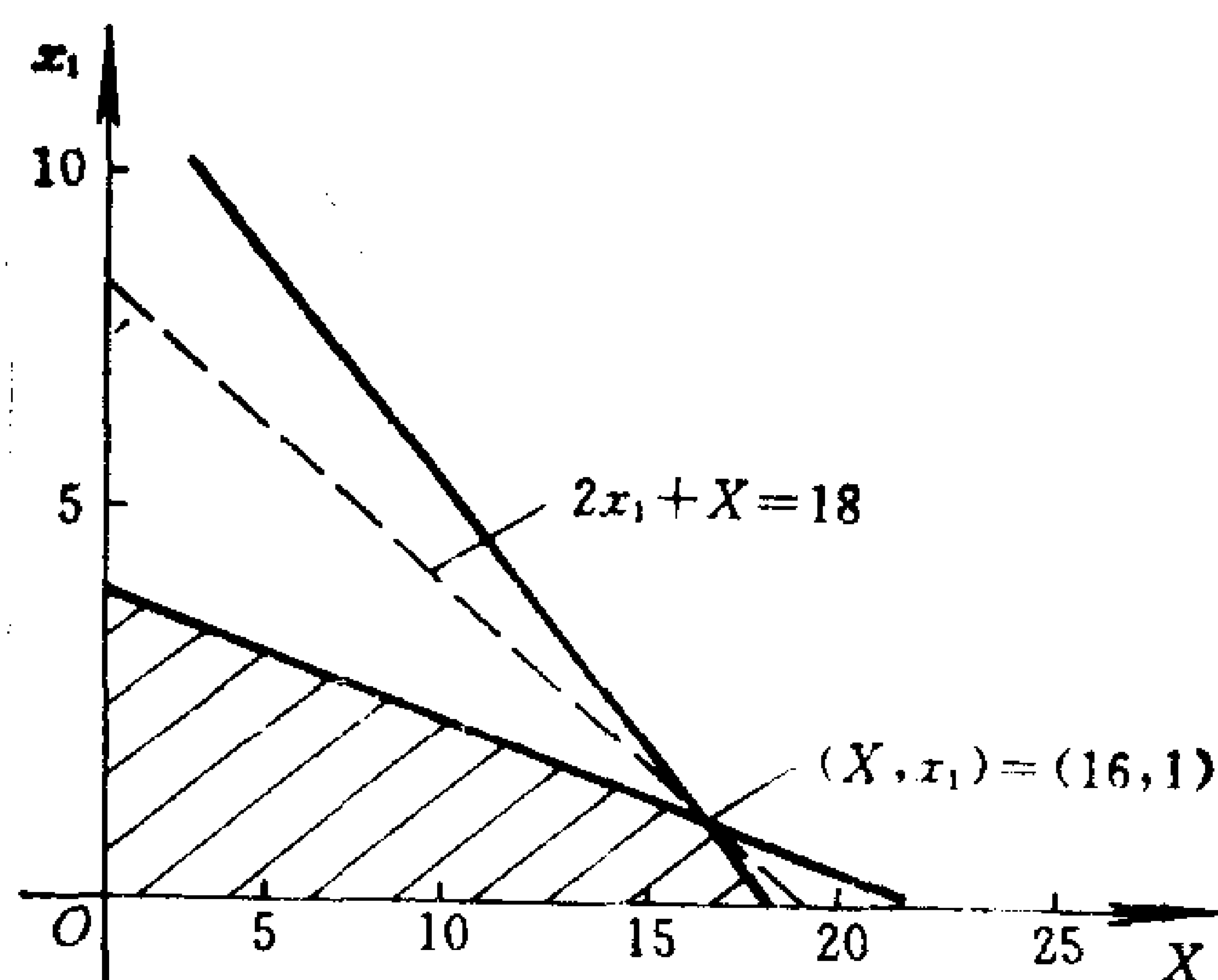


图 1　问题(7)的解

(7) 的图解法见图 1．依 Z_1 增加的方向移动 $2x_1+X$ 等值地通过可行解区域，可以发现在 $(X, x_1)=(16,1)$ 处，$Z_1^*=18$。

剩下的问题是，是否存在(1)的一个可行解，使 $Z=18$？如果存在的话，必定有 $x_1=1$。将它们代入(1)，得不等式组

$$\begin{cases} 2x_2+x_3=16, \\ x_2+x_3\leqslant 11, \\ 2x_2-x_3\leqslant 4, \\ -x_2+x_3\leqslant 1, \\ x_2\geqslant 0, x_3\geqslant 0. \end{cases} \tag{8}$$

解此不等式组，易得

$$5\leqslant x_2\leqslant 5. \tag{9}$$

因此只有 $(x_2, x_3)=(5,6)$ 满足(8)。还可以用图像法解不等式组 (8)，见图 2．给出图 2 的目的主要是表明，最优解不是“凑”出来的。

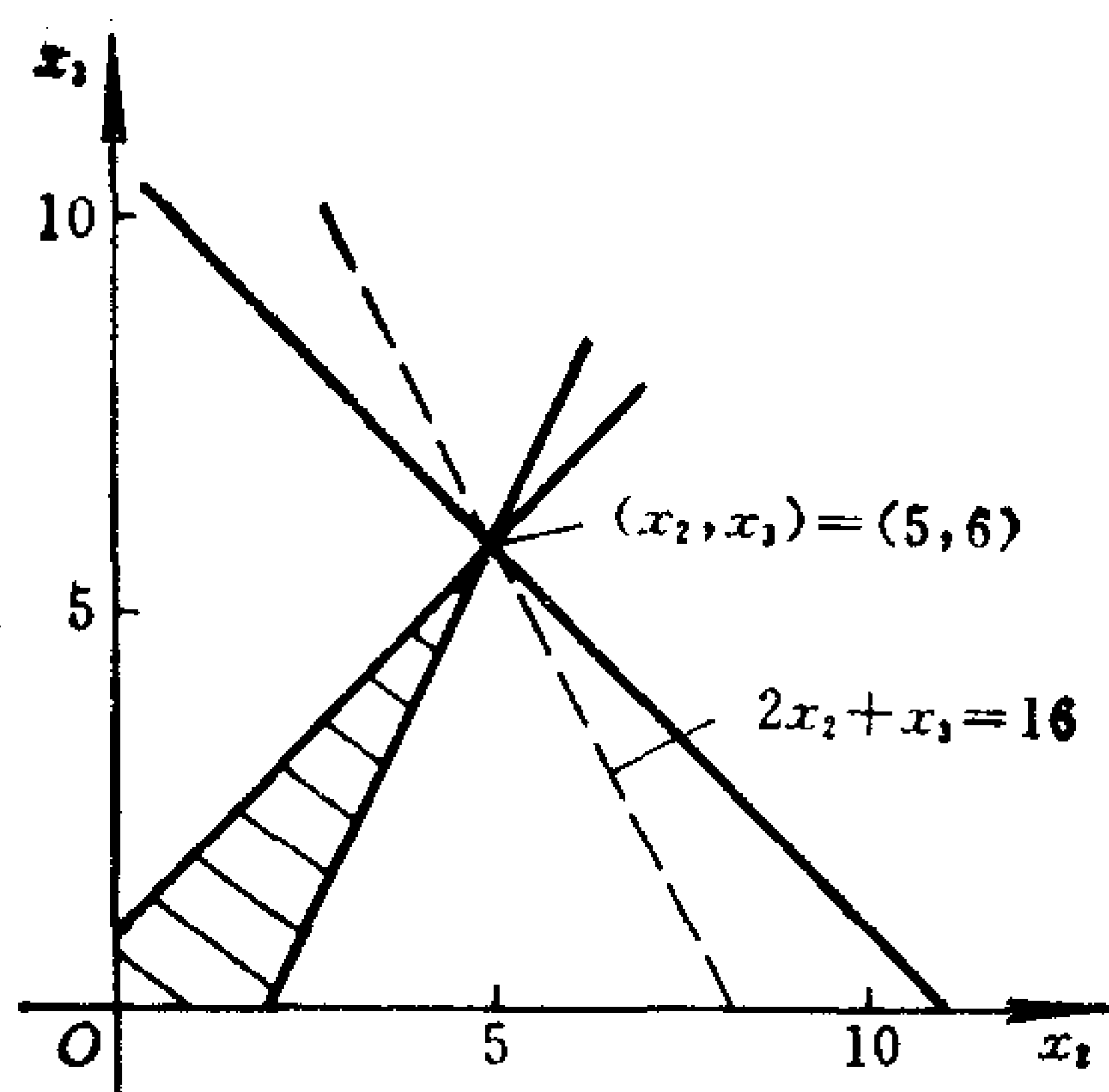


图 2 不等式组(8)的解

现在，问题（1)已完全解出来了，即

$$Z_{max}=18,$$

此时

$$(x_1, x_2, x_3)=(1,5,6)。$$

如果删去解题过程中的讨论，而只考虑图解法本身，则可以发现其中的困难至少不会比用单纯形法求解时遇到的困难大。另外，这个图解法还可以处理含2个以上变量的问题。

3．4个变量的例子

本节给出一个有4个变量及包含3个约束条件的线性规划的例子，它的唯一的最优解也可以用2维图像法得到。

例2　求 $Z=4x_1+5x_2+7x_3-x_4$ 的最大值，其约束条件为

(d) $2x_1 - x_2 + 3x_3 + 4x_4 \leqslant 10$,

(e) $x_1 + x_2 + x_3 - x_4 \leqslant 5$, $x_1, x_2, x_3, x_4 \geqslant 0$. (10)

(f) $x_1 + 2x_2 - 2x_3 + 4x_4 \leqslant 12$,

由于不需要引进什么新概念，因此我们只列出将这一问题化归到能用图像法求解的问题的大致步骤。它们与解方程组时通常使用的步骤相同。首先得到一个简化的含 3 个变量的问题，然后再化简为含 2 个变量的问题，而后者就可以用图像法求解了。如果用审慎的变换得到了所需的简化，就可以再返回去，（利用图像）找到原始问题的解。

选取 $X = 5x_2 + 7x_3$ 作第一次变换。用(d) + k(e)及(d) + k(f)可以得到例2′。

例 2′ 求 $Z_1 = 4x_1 + X - x_4$ 的最大值，其约束条件是

$$\begin{cases} 13x_1 + 2X - 7x_4 \leqslant 65, \\ 35x_1 + 2X + 92x_4 \leqslant 252, \\ x_1 \geqslant 0, X \geqslant 0, x_4 \geqslant 0. \end{cases} \tag{11}$$

继续作变换 $Y = X - x_4$，易得

例 2″ 求 $Z_2 = 4x_1 + Y$ 的最大值，其约束条件是：

$1397x_1 + 198Y \leqslant 7370$， $x_1 \geqslant 0$, Y 可正可负。

用图解法可求得例 2″的解是 $(x_1, Y) = (0, 335/9)$，此时 $Z_2^* = 335/9$。由 $x_1 = 0$，利用图解法可以求得例 2′的解 $(X, x_4) = (352/9, 17/9)$，而 $Z_1^* = Z_2^* = 335/9$。利用 $x_1 = 0$ 及 $x_4 = 17/9$，问题(10)也变成了含两个变量的问题，且它的解是不等式组

$$5x_2 + 7x_3 = \frac{352}{9}, \quad -x_2 + 3x_3 \leqslant \frac{22}{9},$$

$$x_2 + x_3 \leqslant \frac{62}{9}, \quad 2x_2 - 2x_3 \leqslant \frac{40}{9},$$

$$x_2 \geqslant 0, \quad x_3 \geqslant 0$$

的解。可以解得上述不等式组的解是 $(x_2, x_3) = (41/9, 7/3)$，此时 $Z_{max} = 335/9$。因此我们求得了问题(10)的一个可行解，使得 $Z_{max} = Z_1^* = Z_2^* = 335/9$。即，我们得到了最优解$(x_1, x_2, x_3, x_4) = (0, 41/9, 7/3, 17/9)$。

4. 关于可能产生的几种猜测的评注

第一种可能的猜测，产生于对例 1 和例 2 的仔细观察。认为这种技巧只能用于这样一些问题：其中所有“有效”的约束条件只是一些等式。其实并非如此，我们用下面的例子进行说明。

例 3 求 $Z = 4x_1 + 5x_2 - 3x_3$ 的最大值，其约束条件是

$$\begin{aligned} &x_1 + x_2 + x_3 = 10, \quad x_1 - x_2 \geqslant 1, \\ &2x_1 + 3x_2 + x_3 \leqslant 20, \quad x_1 \geqslant 0, x_2 \geqslant 0, x_3 \geqslant 0. \end{aligned} \tag{12}$$

作变换 $X = 4x_1 + 5x_2$ 后得：

求 $Z_1 = X - 3x_3$ 的最大值，约束条件是

$$2X + 9x_3 \leqslant 89, \quad X + 3x_3 \leqslant 40,$$

$$X \geqslant 0, \quad x_3 \geqslant 0.$$

该问题的解是$(X, x_3) = (40, 0)$，此时 $Z_1^* = 40$。因此问题（12）的解是$(x_1, x_2, x_3) = (10, 0, 0)$且 $Z_{max} = 40$。然而，对于这个解，（12）中的第 2 个约束条件并不成为等式。

第 2 个可能的猜测是认为，如果最终能得到只含 2 个变量的线性规划问题且可以用图像法解出，则必有 $Z_{max} = Z_j^*$，其中 j 是化简的步骤数。但这一猜测也不成立。请看下面的

例子.

例 4 求 $Z=10x_1+x_2+2x_3$ 的最大值，约束条件是

$$x_1+x_2-2x_3\leqslant 10,\quad 4x_1+x_2+x_3\leqslant 20,\quad x_1,x_2,x_3\geqslant 0. \tag{13}$$

作变换 $X=10x_1+2x_3$，问题化简为：

求 $Z_1=X+x_2$ 的最大值，约束条件是

$$9X+24x_2\leqslant 460,\quad X\geqslant 0, x_2\geqslant 0. \tag{14}$$

这个问题的解是 $(X,x_2)=(460/9,0)$，此时 $Z_1^*=460/9$. 但可以解得，问题（13）的最优解是 $(x_1,x_2,x_3)=(5,0,0)$ 且 $Z_{max}=50<Z_1^*=460/9$. 另外，还应看到，如果由（14）的解，我们在（13）中替代 $x_2=0$，则可得到它的最优解. 问题是：用这种做法一定能得到 Z_{max} 吗？

可以推得一个显然是正确的推测：如果在最初的约束条件中有一个是等式，则第一次化简总是能够做的. 更令人感兴趣的问题是：如果可以做第一次化简，那么是否一定能得到最终的只含两个变量的简化问题呢？

当然，更大的问题在于：究竟哪类线性规划问题可以用这种技巧求解. 有些问题显然是不行的，例如，起作用的约束条件个数比变量个数至少小 2 的线性规划问题.

结束语 含两个变量的线性规划问题的图解法，按几乎任意的数学水准，都是容易讲授的. 前面给出的几个例子提供了推广这一技巧的方法，并由此产生了一些或易或难解决的猜想，其难易程度视学生的数学水准而定. 但无论如何，这种想法至少是很有趣的.

（朱学贤编译，潘承彪校）

整数的方幂和[①]

B.L.Burrow, R.F.Talbot

1. 引言

Bernoulli 家族中有好几位数学家[②]。第一位是 Jacob (1654—1705)。他的父亲 Nicolaus 是瑞士巴塞尔的商人。据说，Jacob 的座右铭是："不管父亲说什么，我就是研究星星(Invito patre sidera verso)"([1])。不顾父亲的反对，Jacob 毕生致力于数学和天文学的研究。他最著名的工作是《推想的艺术》(Ars Conjectandi')，此书在他去世后于1713年发表。其中包含了他对概率论的兴趣。正是在这本书中，与求整数的方幂和相联系，他引进了著名的 Bernoulli 数。Jacob 热衷于他发现的计算整数方幂和的技巧，并将它与 Bullialdus(1605—1694)的方法进行比较。他说([2])，"借助此表，在不到一刻钟的一半的时间里，我就能算出头 1 千个正整数的10次方相加的和等于

91 409 924 241 424 243 424 241 924 242 500"。

本文将证明一个简单公式，由它可给出这类和的一个高度精确的答案。例如，借助于一个非常简易的计算器，在 9 秒钟

---

① 译自 *The Amer Math. Monthly*, 91(1984), 394—403。

② 参阅《数学译林》第 7 卷第 3 期上的文章"Bernoulli们：一个学者家族"。——译者注

内就可算得

$$\sum_{r=1}^{1000} r^{10} \approx 9.14104 \times 10^{31},$$

它的误差只有0.0005%。

所用的近似公式是

$$\sum_{r=1}^{n} r^k = \frac{(n+1/2)^{k+1}}{k+1}. \tag{1}$$

其思想来自 Bernoulli 的另一个工作：(1) 式左边的和，当k是奇数时等于$p([n(n+1)])$；当n是偶数时等于$(2n+1)\times p([n(n+1)])$，其中$p(x)$是一个多项式([3])。下面我们改进渐近公式(1)，并将结果推广到k是任意实数的情形。

2. $\sum_{r=1}^{n} r^k$(k 是正整数)

熟知这个和可以表成n的$k+1$次多项式。前面提到的 Bernoulli 的工作表明，考虑等价的以$(n+1/2)$为变量的$k+1$次多项式会得到某种好处。

我们用待定系数法求此多项式。记之为$S_{k+1}(n+1/2)$，即

$$\sum_{r=1}^{n} r^k = S_{k+1}(n+1/2), \tag{2}$$

其中

$$S_{k+1}(n+1/2) = \sum_{t=0}^{k+1} a_t (n+1/2)^t. \tag{3}$$

由(2) 式可以推得

$$\sum_{r=1}^{n-1} r^k = S_{k+1}(n-1/2), \tag{4}$$

因而由(2)及(4)式得

$$S_{k+1}(n+1/2) - S_{k+1}(n-1/2) = n^k. \tag{5}$$

以$\binom{k}{r}$表示熟知的二项系数，有

$$(n+1/2)^t - (n-1/2)^t$$

$$= \sum_{j=1}^{t} \binom{t}{j} \left(\frac{1}{2}\right)^j n^{t-j} (1 + (-1)^{j+1}), \tag{6}$$

从而右面或者是 n 的偶数次多项式或者奇数次多项式。等式(5)的右边可改写成

$$\sum_{t=1}^{k+1} a_t \sum_{j=1}^{t} \binom{t}{j} \left(\frac{1}{2}\right)^j n^{t-j} (1 + (-1)^{j+1}) = n^k。 \tag{7}$$

注意系数 a_0 要另外确定。比较恒等式(7)两边的项 $n^p(p=k,k-1,\cdots,0)$ 的系数就可求出系数 $a_t(t=k+1,k,\cdots,1)$.即有

$$\begin{aligned}
a_{k+1}\binom{k+1}{1} &= 1, \\
a_k\binom{k}{1} &= 0, \\
a_{k+1}\binom{k+3}{3}\left(\frac{1}{2}\right)^2 + a_{k-1}\binom{k-1}{1} &= 0, \\
a_k\binom{k}{3}\left(\frac{1}{2}\right)^2 + a_{k-2}\binom{k-2}{1} &= 0,
\end{aligned} \tag{8}$$

及，一般地有

$$a_{k+1}\binom{k+1}{p+1}\left(\frac{1}{2}\right)^{p+1}(1+(-1)^{p+2})+a_k\binom{k}{p}\left(\frac{1}{2}\right)^{p}$$

$$\times(1+(-1)^{p+1})+a_{k-1}\binom{k-1}{p-1}\left(\frac{1}{2}\right)^{p-1}(1+(-1)^{p})$$

$$+\cdots+a_{k+1-p}\binom{k+1-p}{1}=0.$$

首先可以看到，由上述等式中的第 2 个得 $a_k=0$，由此得 a_{k-2}，a_{k-4}，…，均为零。由第 1 个等式得 a_{k+1}，然后依次可解出 a_{k-1}，a_{k-3}，…，我们有

$$a_{k+1}=\frac{1}{k+1},$$

$$a_{k-1}=-\frac{1}{24}\binom{k}{1},$$

$$a_{k-3}=\frac{7}{960}\binom{k}{3}, \tag{9}$$

$$a_{k-5}=\frac{-31}{2^7\cdot 63}\binom{k}{5},$$

$$a_{k-7}=\frac{127}{2^7\cdot 240}\binom{k}{7}.$$

例如 $k=3$ 时得

$$\sum_{r=1}^{n}r^3=S_4(n+1/2)=\frac{1}{4}(n+1/2)^4-\frac{1}{8}(n-1/2)^2+a_0.$$

当 $n=1$ 时 $S_4(3/2)=1$，从而 $a_0=1/64$。

因此

$$S_4(n+1/2)=\frac{1}{4}\{(n+1/2)^2-1/4\}^2$$

$$=\frac{1}{4}n^2(n+1)^2. \tag{10}$$

最后的形式是大家熟知的。

从计算的角度看上述结果也许更有价值。记$n+1/2$为x，$\sum_{r=1}^{n} r^k$ 为 $S(k)$，我们已经证明了

$$S(k)=a_{k+1}x^{k+1}+a_{k-1}x^{k-1}+a_{k-3}x^{k-3}+\cdots, \tag{11}$$

其中的系数由(9)式给出。而且当 k 是奇数时，最后一项是 a_0；当 k 是偶数时，最后一项是 a_1x，于是，对于$k\geqslant4$，(11)式可写成

$$S(k)=\frac{x^{k+1}}{k+1}\left\{1-\frac{1}{12x^2}\binom{k+1}{2}+\frac{7}{240x^4}\binom{k+1}{4}-\cdots\right\}. \tag{12}$$

显然，由(12)式可以看到，由近似公式

$$S_k\approx\frac{x^{k+1}}{k+1} \tag{13}$$

可以得到很好的计算结果。为说明这些结果，考虑$\sum_{r=1}^{35} r^5$，将其精确值333,263,700 与由(11)(等价地，由(12)) 所算得的两个近似值作比较。

第一个近似值由(13)给出

$$\sum_{r=1}^{35} r^5\approx\frac{35.5^6}{6}=333\,594\,489.$$

它的误差约0.1%。

取(11)(或(12))式的前两项得第2个近似值，即

$$\sum_{r=1}^{35} r^5 \approx \frac{35.5^6}{6} - \frac{5}{24}(35.5)^4 = 333\,263\,608.$$

其绝对误差是93，相对误差约为0.0000003%.

上述工作的自然推广是讨论k不是整数的情形。

3. 一般情形

本节讨论和$\sum_{r=1}^{n} r^k$，其中的k是实数。我们将导出一个结果，是上节中公式(12)的推广。特别地，我们将证明：当k为正数时，在(13)中得到的简单近似公式给出令人吃惊的精确估计。

当k不是正整数时，下面导出的公式来自关于整数方幂和的渐近分析方法的一个应用。有关这种技巧的讨论可在[4]中找到。特别地，利用n的方幂及Riemann-Zeta函数，[4]中的等式3.4.6给出了有关收敛情形的一个结果。

考虑$f(x)=x^k$在整数$x=r$处的Taylor 展开式

$$f(x)=f(r)+(x-r)f'(r)+\frac{(x-r)^2}{2!}f''(r)+\cdots$$

$$+\frac{(x-r)^q}{q!}f^{(q)}(r)+\frac{(x-r)^{q+1}}{(q+1)!}f^{(q+1)}(z_r), \tag{14}$$

其中$z_r=r+\theta_r(x-r)$，$0<\theta_r<1$。

对非整数k及整数r，利用记号

$$\binom{k}{r}=\frac{k(k-1)\cdots(k-r+1)}{r!} \tag{15}$$

可以将(14)式写成如下的形式

$$x^k=r^k+\binom{k}{1}r^{k-1}(x-r)+\binom{k}{2}r^{k-2}(x-r)^2+\cdots$$

$$+\binom{k}{q}r^{k-q}(x-r)^q+\binom{k}{q+1}z_r^{k-q-1}(x-r)^{q+1},\tag{16}$$

不失一般性，可以取其中的 q 为偶数。上式两边从$r-1/2$到$r+1/2$积分得

$$\int_{r-1/2}^{r+1/2}x^k\mathrm{d}x=r^k+\binom{k}{2}r^{k-2}\cdot\frac{2}{3}\left(\frac{1}{2}\right)^3$$

$$+\binom{k}{4}r^{k-4}\cdot\frac{2}{5}\left(\frac{1}{2}\right)^5+\cdots.\tag{17}$$

从$r=1$到$r=n$对等式(17)求和得

$$g(k)=\sum_{r=1}^{n}r^k+\binom{k}{2}r^{k-2}\frac{1}{3}\left(\frac{1}{2}\right)^2\sum_{r=1}^{n}r^{k-2}$$

$$+\binom{k}{4}\frac{1}{5}\left(\frac{1}{2}\right)^4\sum_{r=1}^{n}r^{k-4}+\cdots,\tag{18}$$

其中

$$g(k)=\frac{1}{k+1}\left\{\left(n+\frac{1}{2}\right)^{k+1}-\left(\frac{1}{2}\right)^{k+1}\right\},\qquad k\neq-1,$$

及$g(-1)=\log(2n+1)$。记$S(k)=\sum_{r=1}^{n}r^k$，等式(18)可改写成

$$g(k)=S(k)+\binom{k}{2}\frac{1}{3\cdot 2^2}S(k-2)+\binom{k}{4}\frac{1}{5\cdot 2^4}S(k-4)+\cdots$$

$$+\binom{k}{q}\frac{1}{(q+1)\cdot 2^{q}}S(k-q)+R_{q+1}, \tag{19}$$

其中

$$R_{q+1}=\binom{k}{q+1}\sum_{r=1}^{n}\int_{r-1/2}^{r+1/2}z_r^{k-q-1}(x-r)^{q+1}\mathrm{d}x,$$

而z_r由(14)式给出。

于是，对$q>k-1$有

$$|R_{q+1}|<\left|\binom{k}{q+1}\right|\sum_{r=1}^{n}\int_{r-1/2}^{r+1/2}\left(r-\frac{1}{2}\right)^{k-q-1}|x-r|^{q+1}\mathrm{d}x,$$

即

$$|R_{q+1}|<\left|\binom{k}{q+1}\right|\sum_{r=1}^{n}\left(r-\frac{1}{2}\right)^{k-q-1}\frac{2(1/2)^{q+2}}{q+2}, \tag{20}$$

因此有

$$|R_{q+1}|<\left|\binom{k}{q+1}\right|\frac{n}{q+2}\left(\frac{1}{2}\right)^{k}.$$

对固定的$k\geqslant -1$及 n，由此容易推得

$$\lim_{q\to\infty}R_{q+1}=0,$$

因此级数(19)收敛。

在(19)式中依次取$k,k-2,k-4,\cdots$，得到一组等式

$$g(k)=S(k)+\binom{k}{2}\frac{1}{3\cdot 2^{2}}S(k-2)+\binom{k}{4}\frac{1}{5\cdot 2^{4}}S(k-4)$$

$$+\binom{k}{6}\frac{1}{7\cdot 2^{6}}S(k-6)+\cdots,$$

$$g(k-2)=S(k-2)+\binom{k-2}{2}\frac{1}{3\cdot 2^{2}}S(k-4)$$

$$+\binom{k-2}{4}\frac{1}{5\cdot 2^4}S(k-6)+\cdots,$$

$$g(k-4)=S(k-4)+\binom{k-4}{2}\frac{1}{3\cdot 2^2}S(k-6)+\cdots,$$

$$g(k-6)=S(k-6)+\cdots,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \tag{21}$$

用矩阵形式写出得

$$g=Ms, \tag{22}$$

其中

$$g=\begin{pmatrix} g(k) \\ g(k-2) \\ g(k-4) \\ \vdots \end{pmatrix}, \qquad s=\begin{pmatrix} S(k) \\ S(k-2) \\ S(k-4) \\ \vdots \end{pmatrix}$$

及

$$M=\begin{pmatrix} 1 & \binom{k}{2}\frac{1}{3\cdot 2^2} & \binom{k}{4}\frac{1}{5\cdot 2^4} & \binom{k}{6}\frac{1}{7\cdot 2^6} & \cdots \\ 0 & 1 & \binom{k-2}{2}\frac{1}{3\cdot 2^2} & \binom{k-2}{4}\frac{1}{5\cdot 2^4} & \cdots \\ 0 & 0 & 1 & \binom{k-4}{2}\frac{1}{3\cdot 2^2} & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

我们着手解决问题的方法是从无穷矩阵M中截取 l 行 l 列的有穷矩阵 M_l，类似地，从向量g中截取 g_l，并定义 s_l 使 $M_l s_l = g_l$。

值得注意的是，矩阵M具有这样的性质：若对给定的 l 求得逆矩阵M_l^{-1}，那么这个M_l^{-1}是更高阶逆矩阵的子矩阵。作为特例，取$l=4$，我们有

$$M_4^{-1}=\begin{pmatrix} 1 & -\binom{k}{2}\frac{1}{3\cdot 2^2} & \binom{k}{4}\frac{7}{15\cdot 2^4} & -\binom{k}{6}\frac{31}{21\cdot 2^6} \\ 0 & 1 & -\binom{k-2}{2}\frac{1}{3\cdot 2^2} & \binom{k-2}{4}\frac{7}{15\cdot 2^4} \\ 0 & 0 & 1 & -\binom{k-4}{2}\frac{1}{3\cdot 2^2} \\ 0 & 0 & 0 & 1 \end{pmatrix},\tag{23}$$

因此有

$$s_4=M_4^{-1}g_4.\tag{24}$$

由此可得 $S(k)$的由 4 项组成的近似公式

$$S_4(k)=g(k)-\binom{k}{2}\frac{1}{3\cdot 2^2}g(k-2)+\binom{k}{4}\frac{7}{15\cdot 2^4}g(k-4)-\binom{k}{6}\frac{31}{21\cdot 2^6}g(k-6).\tag{25}$$

当 k 为正整数时，(25)式与前面的结果((9)及(11)式)是一致的。

正如从下面的例子可以看到的那样，取头几项，就显示了这一结果在数值上的精确性。我们有

$$\sum_{r=1}^{35}r^{2.5}=76\,138.722369.$$

而下表给出当 l 取不同数值时的近似值。应该指出，近似值

l	近 似 值
1	76160 721097
2	76138.725044
3	76138.722204
4	76138.722459

$(35.5)^{3.5}/3.5$——它与近似公式(13)类似——的误差仅仅只有约0.003%。

本节中所用的矩阵技巧也能用于 k 是正整数的情形。因为这时 M 是 m 阶有限矩阵，所以只要假定所限定的阶 l 不超过 m，前面的分析仍成立。

4. 当k是负数时的修正

实际上，当k是负数时，对等式(17)的求和若从 $r=p$ 开始，就可以提高公式(25)的精确性。这等价于把 $S(k)$的定义改为

$$S(k)=\sum_{r=p}^{n} r^{k}, \tag{26}$$

且相应地定义

$$g(k)=\frac{1}{k+1}\left\{\left(n+\frac{1}{2}\right)^{k+1}-\left(p-\frac{1}{2}\right)^{k+1}\right\}, \quad k\neq -1, \tag{27}$$

及

$$g(-1)=\log(n+1/2)-\log(p-1/2). \tag{28}$$

例如，当$k=-1$时，有明确的形式

$$\sum_{r=p}^{n}\frac{1}{r}=\left\{\log\left(n+\frac{1}{2}\right)-\log\left(p-\frac{1}{2}\right)\right\}$$

$$+\frac{1}{24}\left\{\left(n+\frac{1}{2}\right)^{-2}-\left(p-\frac{1}{2}\right)^{-2}\right\}$$

$$-\frac{7}{960}\left\{\left(n+\frac{1}{2}\right)^{-4}-\left(p-\frac{1}{2}\right)^{-4}\right\}$$

$$+\frac{31}{63\cdot 2^{7}}\left\{\left(n+\frac{1}{2}\right)^{-6}-\left(p-\frac{1}{2}\right)^{-6}\right\}-\cdots. \qquad (29)$$

这样，对 $p=3$及 $n=50$，利用前面说过的逆矩阵技巧，计算一项得近似值3.0056826；计算五项可得近似值2.9992059。它可以与精确到 7 位小数的近似值2.9992053相比较。

将熟知的渐近公式

$$\sum_{r=1}^{n}\frac{1}{r}\approx\log n+\gamma \qquad (30)$$

(其中的$\gamma=0.57721566\cdots$是 Euler 常数)与我们这里得到的第一个渐近公式

$$\sum_{r=1}^{n}\frac{1}{r}\approx\sum_{r=1}^{p-1}\frac{1}{r}+\log\frac{n+1/2}{p-1/2} \qquad (31)$$

及第二个渐近公式

$$\sum_{r=1}^{n}\frac{1}{r}\approx\sum_{r=1}^{p-1}\frac{1}{r}+\log\frac{n+1/2}{p-1/2}$$

$$+\frac{1}{24}\left\{\left(n+\frac{1}{2}\right)^{-2}-\left(p-\frac{1}{2}\right)^{-2}\right\} \qquad (32)$$

相比较也是很有趣的。对$n=50$，由(30)式得近似值4.48924，它与精确值4.499205相比较误差只有0.22%。下表给出对不

同的 p 值由近似公式(31)及(32)得到的近似值：

近似公式	$p=2$	$p=3$	$p=4$
(31)	4.51651	4.50568	4.50254
(32)	4.49801	4.49903	4.49916

至此我们只讨论了形如 $\sum_{r=1}^{n} r^k$ 的 有限级数的例子，因为当$k\geqslant -1$及 $n\to+\infty$时这个和不存在。但当 $k<-1$时，可以利用(25)及(27)式估算$\sum_{r=1}^{\infty} r^k$。这时，令$n\to\infty$得

$$\sum_{r=1}^{\infty} r^k=\sum_{r=1}^{p-1} r^k-\frac{1}{k+1}\left(p-\frac{1}{2}\right)^{k+1}+\frac{k}{24}\left(p-\frac{1}{2}\right)^{k-1}$$

$$-\binom{k}{3}\frac{7}{960}\left(p-\frac{1}{2}\right)^{k-3}$$

$$+\binom{k}{5}\frac{31}{8064}\left(p-\frac{1}{2}\right)^{k-5}-\cdots. \tag{33}$$

容易修改(19)式的收敛性证明，使之适用于 n 趋于无穷及$k<-1$的情形。从相应于现在所讨论的情形的(20)式，可得

$$|R_q|<\frac{1}{2^q(q+1)}\binom{k}{q}\sum_{r=p}^{\infty}\left(r-\frac{1}{2}\right)^{k-q}.$$

因为$k<-1$，上式中的无穷和收敛，所以

$$\lim_{q\to\infty} R_q=0.$$

利用(33)式中给出的这些项，对$k=-2$，$k=-3.5$及不同的

p 值计算无穷级数的值，得

k	$p=2$	$p=3$	$p=4$	精确值
-2	1.644466	1.644928	1.644934	1.644934
-3.5	1.124871	1.126723	1.126733	1.126730

5. 收敛性

在前几节中，我们通过取级数

$$\sum_{l=0}^{\infty} m_l g(k-2l) \tag{34}$$

的头几项来计算 $\sum_{r=p}^{n} r^k$ 的近似值，其中的 m_l 是逆矩阵 M_l^{-1} 的第 1 行中的第 l 项（见(22)—(25)式），$g(k)$ 由(27)式给出。

事实上，(34)式仅当 k 是正整数时收敛，而实际上此时级数是一有限和。在其余所有的情形级数(34)均发散。下面我们证明，级数一般项的绝对值是递增的。

首先，由(27)式得

$$g(k-2l)=\frac{1}{k-2l+1}\left\{\left(n+\frac{1}{2}\right)^{k-2l+1}-\left(p-\frac{1}{2}\right)^{k-2l+1}\right\},$$

因此对充分大的 l 及 n 有

$$|g(k-2l)|>\frac{1}{|k-2l+1|}\cdot\frac{1}{p^{2l-k-1}}, \quad p\geqslant 1. \tag{35}$$

矩阵 M_l 可以写成形式

$$
M_l=\begin{pmatrix}
1 & a(k) & \lambda_1 a(k)a(k-2) & \lambda_2\prod\limits_{i=0}^{2}a(k-2i) & \cdots & \lambda_{l-2}\prod\limits_{i=0}^{l-2}a(k-2i)\\
0 & 1 & a(k-2) & \lambda_1 a(k-2)a(k-4) & \cdots & \lambda_{l-3}\prod\limits_{i=1}^{l-2}a(k-2i)\\
0 & 0 & 1 & a(k-4) & \cdots & \lambda_{l-4}\prod\limits_{i=2}^{l-2}a(k-2i)\\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\
0 & 0 & 0 & 0 & \cdots & 1
\end{pmatrix},
$$

其中

$$
a(k)=\binom{k}{2}\frac{1}{3\cdot 2^2},\qquad \lambda_r=\frac{6^{r+1}}{(2r+3)!}.
$$

考虑方程

$$
M_l x=e, \tag{36}
$$

其中 e 是 $l\times l$ 阶单位矩阵的第 l 列。设 x 的第 1 个分量 $x_1=m_l$。利用 Cramer 法则并注意到 M_l 的行列式等于 1，我们有

$$
m_l=\begin{vmatrix}
0 & a(k) & \lambda_1 a(k)a(k-2) & \lambda_2\prod\limits_{i=0}^{2}a(k-2i) & \cdots & \lambda_{l-2}\prod\limits_{i=0}^{l-2}a(k-2i)\\
0 & 1 & a(k-2) & \lambda_1 a(k-2)a(k-4) & \cdots & \cdots\\
0 & 0 & \cdots & \cdots & \cdots & \cdots\\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\
1 & 0 & \cdots & \cdots & \cdots & 1
\end{vmatrix}
$$

$$
=(-1)^{l+1}\begin{vmatrix}
a(k) & \lambda_1 a(k)a(k-2) & \lambda_2\prod\limits_{i=0}^{2}a(k-2i) & \cdots & \lambda_{l-2}\prod\limits_{i=0}^{l-2}a(k-2i)\\
1 & a(k-2) & \lambda_1 a(k-2)a(k-4) & \cdots & \cdots\\
0 & 1 & a(k-4) & \cdots & \cdots\\
\cdots & \cdots & \cdots & \cdots & \cdots\\
0 & 0 & 0 & \cdots & a(k-2l+4)
\end{vmatrix}.
\tag{37}
$$

对 l 用归纳法可以证明

$$m_l = (-1)^{l+1} a(k)a(k-2)\cdots a(k-2l+4)\Delta_l, \tag{38}$$

其中

$$\Delta_l = \begin{vmatrix} 1 & \lambda_1 & \lambda_2 & \cdots & \lambda_{l-2} \\ 1 & 1 & \lambda_1 & \cdots & \lambda_{l-3} \\ 0 & 1 & 1 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \lambda_1 \\ 0 & 0 & 0 & \cdots & 1 \end{vmatrix}。 \tag{39}$$

相继消去主对角线下面的那些1（只要从第2行开始，每行减去前一行的若干倍就行），就可以求得Δ_l的值。这时Δ_l成为一个上三角形行列式

$$\Delta_l = \begin{vmatrix} 1 & \lambda_1 & \lambda_2 & \cdots & \lambda_{l-2} \\ 0 & f_1 & \cdots & \cdots & \cdots \\ 0 & 0 & f_2 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & f_{l-2} \end{vmatrix}, \tag{40}$$

从而 $\Delta_l = f_1 f_2 \cdots f_{l-2}$。当 l 增大时前面的那些 f_i 值不会改变。例如

$$\Delta_{l+1} = f_1 f_2 \cdots f_{l-2} f_{l-1}。 \tag{41}$$

具体计算数值表明数列$\{f_i\}$是单调递减的，且迅速趋于某一极限。下表给出了f_i的一些值。

定义数列$\{f_i\}$的极限是 $1/\beta(i\to\infty)$，可见当 $l\to\infty$ 时有$\Delta_l\to 0$。另外，由单调性可知

$$\Delta_l > (1/\beta)^{l-1}, \tag{42}$$

i	f_i
1	0.7
2	0.63265306
13	0.60792713
14	0.60792711
15	0.60792710
16	0.60792710

而且$\prod_{i=0}^{l-2} a(k-2i)$可写为

$$k(k-1)\cdots(k-2l+3)/24^{l-1}. \tag{43}$$

于是，综合本节所得的结果，我们证明了

$$|m_l g(k-2l)| > \frac{|k(k-1)\cdots(k-2l+3)|}{(24\beta)^{l-1}}\left|\frac{p^{k-2l+1}}{k-2l+1}\right|. \tag{44}$$

因为p和k是固定的，所以(44)式表明级数(34)的项的绝对值最终是不断增加的。从而对充分大的l，这些近似公式将变得很糟糕。那么，本文中给出的估算为什么是好的呢？考察方程组(21)的截段形式

$$M_l \hat{S}_l = g_l + \varepsilon_l, \tag{45}$$

其中$\hat{S}_l$是由前l个精确解组成的列向量，$-\varepsilon_l$是前l个方程的截断误差(见(20)式)组成的列向量。这样，$S(k)$的精确表达式是由形如

$$m_t(g(k-2t)+\eta_t) \tag{46}$$

的项组成的有限和，其中η_t是向量ε_t的元素。取特殊情形$k=3$得

$$m_3=\frac{7k(k-1)(k-2)(k-3)}{4!\cdot 15\cdot 2^4}. \tag{47}$$

因为η_3很小，而且对于正的k这些和的阶为$(n+1/2)^{k+1}/(k+1)$，所以对n充分大，相对误差

$$|m_3\eta_3|/S(k)$$

很小。对负的k(特别地，对$k<-1$)，近似公式是利用$p>1$来构造的，因为η_3随p增大而减小，所以总有可能去缩小$|m_3\eta_3|$。当n或者p充分大时，这些论证对任意大都成立，因而有可能取级数(34)的头几项得到一个好的估算。这样，对于为什么能从级数(34)的一些项得到$S(k)$的令人吃惊地好的近似值，我们给出了一个启发性的说明，但级数(34)本身确确实实是发散的！

参 考 文 献

[1] E. T. Bell, Men of Mathematics, 1937.

[2] D. E. Smith, A Source Book in Mathematics, 1929.

[3] A. W. F. Edwards, Sum of powers of integers: a little of the history, *Math. Gaz.* (1982), 22—28.

[4] N. G. de Brujn, Asymptotic Methods in Analysis, 1958.

（潘承彪译，朱学贤校）

方幂和的快速计算法①

A.W.F.Edwards

一个令人感兴趣的事实是：正整数的 r 阶方幂和

$$\sum_{\nu=1}^{n}\nu^{r}=a_1n+a_2n^2+\cdots+a_{r+1}n^{r+1} \qquad (1)$$

可以用 1 阶方幂和

$$\sum_{\nu=1}^{n}\nu=n(n+1)/2$$

及 2 阶方幂和

$$\sum_{\nu=1}^{n}\nu^2=n(n+1)(2n+1)/6$$

表示出来。这一结果最终可追溯到 Bernoulli 多项式的对称性。它的第一个例子自然是大家熟悉但又感到惊奇的关系式

$$\sum_{\nu=1}^{n}\nu^3=\left(\sum_{\nu=1}^{n}\nu\right)^2.$$

一般性的结果由 Jacobi([3])于1834年证得。

但是，也许比结果本身更令人诧异的是：在Jacobi 的工作发表之前 200 年，有人就已知晓了这一结果。1981年，当

① A quick route to sums of powers, *Amer. Math. Monthly*, 93(1986), 451—455.

我在研究 James Bernoulli 的著作《推想的艺术》(Ars Conjectandi, Basel, 1713)时，发现一条线索，一位当今已被人遗忘的德国数学家Johann Faulhaber在其所著的《学院代数》(Academia Algebrae, Augspurg, 1631)一书中已叙述了这一结果。Bernoulli在引进多项式(1)并列出系数表(其中也包括了de Moivre称谓的Bernoulli数)时，也提到了Faulhaber的名字。非常巧的是，我很顺利地得到了Faulhaber 的这本书，它是剑桥大学收藏的《学院代数》的抄本，而且一度曾是Jacobi的藏书(不过，不能确定Jacobi究竟是在1834年之前还是在之后得到它的)。

为书写简便，下文记$\sum_{\nu=1}^{n}\nu^r$为$\sum n^r$。可以发现，Faulhaber的多项式是：

(i) 当 r 是偶数时，

$$\sum n^r=\sum n^2\cdot\left(b_1+b_2\sum n+b_3\left(\sum n\right)^2+\cdots+b_{r/2}\left(\sum n\right)^{r/2-1}\right);$$

(ii) 当$r\geqslant 3$是奇数时

$$\sum n^r=\left(\sum n\right)^2\left(c_1+c_2\sum n+c_3\left(\sum n\right)^2+\cdots+c_{(r-1)/2}\left(\sum n\right)^{(r-3)/2}\right),\qquad(2)$$

其中的系数b_i及c_i与 r 有关。

Faulhaber 给出了一种算法，对于预先给定的 r，可由系数 b_i 得到系数 c_i，同时也给出了计算b_i 的方法。我给(2)

式起名为“Faulhaber 多项式”([1])。Schneider([4])也提到过Faulhaber 的方法。

本文的目的是展示 Faulhaber 多项式的矩阵形式，所用的方法类似于多项式(1)的矩阵表示([1]).这样的处理方式，不仅显得非常优美，而且还使得Faulhaber的从b_i得到c_i的算法更易于被理解。

如同 Tits([5])所做的那样，考虑$[x(x+1)]^r-[x(x-1)]^r$的展开式，并采用 Pascal 的方法，对$x=1,2,\cdots,n$依次写出等式然后求和。我们得到$\left(\text{符号}\sum\text{的意义同前}\right)$

$$[n(n+1)]^r=2\left[r\sum n^{2r-1}+\binom{r}{3}\sum n^{2r-3}+\binom{r}{5}\sum n^{2r-5}+\cdots\right]. \tag{3}$$

写成矩阵形式得

$$\begin{pmatrix}[n(n+1)]^2\\ [n(n+1)]^3\\ [n(n+1)]^4\\ [n(n+1)]^5\\ \cdots\end{pmatrix}=2\begin{pmatrix}2 & & & & \\ 1 & 3 & & 0 & \\ 0 & 4 & 4 & & \\ 0 & 1 & 10 & 5 & \\ \cdots & \cdots & \cdots & \cdots & \end{pmatrix}\begin{pmatrix}\sum n^3\\ \sum n^5\\ \sum n^7\\ \sum n^9\\ \cdots\end{pmatrix}, \tag{4}$$

其中的行对应于$r=2,3,4,\cdots$。(4)式中间的那个矩阵的每一

行，是由Pascal 三角形的相应行一隔一地删去元素而得到的（即，对于Pascal三角形中的奇数行，删去其偶数项系数，而偶数行则删去其奇数项系数——译注）。记 $u=n(n+1)$。解方程组(4)，对于奇次方幂和得

$$\begin{pmatrix} \sum n^3 \\ \sum n^5 \\ \sum n^7 \\ \sum n^9 \\ \cdots \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 2 & & & & \\ 1 & 3 & & 0 & \\ 0 & 4 & 4 & & \\ 0 & 1 & 10 & 5 & \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}^{-1}\begin{pmatrix} u^2 \\ u^3 \\ u^4 \\ u^5 \\ \cdots \end{pmatrix}, \tag{5}$$

因为 $u=2\sum n$，所以这是“奇次” Faulhaber 多项式的完全解。特别地，u^2 是每一个多项式的因子，因此这证明了形式(2)。

设(4)式中的矩阵为 $F=\{f_{ij}\}$，则有

$$f_{ij}=\binom{i+1}{2(i-j)+1} \tag{6}$$

或者 $f_{ij}=0$，如果对于 i 及 j，(6)式不定义二项式系数（例如，$f_{11}=\binom{2}{1}=2$，而当 $j\geqslant 2$ 时 $\binom{2}{2(1-j)+1}$ 不定义二项式系数，因此 $f_{ij}=0$——译注）。逆矩阵

$$F^{-1}=\begin{pmatrix} 1/2 & & & \\ -1/6 & 1/3 & & 0 \\ 1/6 & -1/3 & 1/4 & \\ -3/10 & 3/5 & -1/2 & 1/5 \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}, \qquad (7)$$

注意到$u=2\sum n$中的因子 2，这就导出了奇 次 Faulhaber 多项式中的所有系数。

Tits还考虑了偶次多项式。他用 Pascal 的方法于 $x^r(x+1)^{r+1}-x^{r+1}(x-1)^r$的展开式，并用(3)式去掉 奇 次幂。写成矩阵形式。这一结果即为

$$\begin{pmatrix} \sum n^2 \\ \sum n^4 \\ \sum n^6 \\ \sum n^8 \\ \cdots \end{pmatrix}=\frac{1}{2}(2n+1)\begin{pmatrix} 3 & & & \\ 1 & 5 & & 0 \\ 0 & 5 & 7 & \\ 0 & 1 & 14 & 9 \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}^{-1}\begin{pmatrix} u \\ u^2 \\ u^3 \\ u^4 \\ \cdots \end{pmatrix}. \qquad (8)$$

记(8)中的矩阵为$G=\{g_{ij}\}$，则有

$$g_{ij}=\binom{i+1}{2(i-j)+1}+\binom{i}{2(i-j)+1}, \qquad (9)$$

其中，无定义的二项式系数仍用0来代替。可以看到

$$G = F + \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0 & & & & \\ 0 & & F & & \\ 0 & & & & \\ \cdots & & & & \end{pmatrix}, \tag{10}$$

于是，由逆矩阵

$$G^{-1} = \begin{pmatrix} 1/3 & & & & \\ -1/15 & 1/5 & & 0 & \\ 1/21 & -1/7 & 1/7 & & \\ -1/15 & 1/5 & -2/9 & 1/9 & \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{11}$$

就可导出偶次 Faulhaber 多项式的所有系数。

Faulhaber 本质上是知道如何由 G^{-1} 去求 得 F^{-1} 的。我们将他的算法演示如下。

取 F，并将其每一行的元素除以该行中的对角线元，得

$$F = \begin{pmatrix} 2 & & & & \\ 1 & 3 & & 0 & \\ 0 & 4 & 4 & & \\ 0 & 1 & 10 & 5 & \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

$$= \begin{pmatrix} 2 & & & & \\ & 3 & & 0 & \\ & & 4 & & \\ & & & 5 & \\ & 0 & & & \ddots \end{pmatrix} \begin{pmatrix} 1 & & & & \\ 1/3 & 1 & & 0 & \\ 0 & 1 & 1 & & \\ 0 & 1/5 & 2 & 1 & \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{12}$$

再取 G，并将每一列的元素除以该列中的对角线元，得

$$G=\begin{pmatrix} 3 & & & & \\ 1 & 5 & & 0 & \\ 0 & 5 & 7 & & \\ 0 & 1 & 14 & 9 & \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

$$=\begin{pmatrix} 1 & & & & \\ 1/3 & 1 & & 0 & \\ 0 & 1 & 1 & & \\ 0 & 1/5 & 2 & 1 & \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}\begin{pmatrix} 3 & & & & 0 \\ & 5 & & & \\ & & 7 & & \\ & & & 9 & \\ 0 & & & & \ddots \end{pmatrix} \tag{13}$$

比较(12)和(13)可以发现，上面导出的矩阵中有两个矩阵是相同的。当用(6)和(9)来分析时可知，这个相等关系是建立在关于二项式系数的一个简单的（并非重要的）等式之上的。

记

$$X=\begin{pmatrix} 2 & & & & 0 \\ & 3 & & & \\ & & 4 & & \\ 0 & & & 5 & \\ & & & & \ddots \end{pmatrix}$$

及

$$Y=\begin{pmatrix} 3 & & & & 0 \\ & 5 & & & \\ & & 7 & & \\ 0 & & & 9 & \\ & & & & \ddots \end{pmatrix},$$

则有

$$X^{-1}F=GY^{-1},$$

或者取逆得

$$F^{-1}X=YG^{-1}.$$

从而有

$$F^{-1}=YG^{-1}X^{-1}. \tag{14}$$

容易看到，用Y左乘G^{-1}，然后再右乘X^{-1}，相当于用3,5,7,9,…,分别去乘G^{-1}的行，再用2,3,4,5,…，分别去除其列。因此，为了得到F^{-1}的第i行第j列的元素，我们只要取G^{-1}中的相应元素，乘以$(2i+1)$再除以$(j+1)$即可。

例如，当$i=4$时，G^{-1}中的第4行中的前4个元素是

$$-1/15,\ 1/5,\ -2/9,\ 1/9.$$

分别乘上

$$9/2,\ 9/3,\ 9/4,\ 9/5$$

后得到F^{-1}的第4行的前4个元素为

$$-3/10,\ 3/5,\ -1/2,\ 1/5.$$

因此，由多项式

$$\sum n^8=\frac{1}{2}(2n+1)\left(-\frac{1}{15}u+\frac{1}{5}u^2-\frac{2}{9}u^3+\frac{1}{9}u^4\right)$$

导得

$$\sum n^9=\frac{1}{2}\left(-\frac{3}{10}u^2+\frac{3}{5}u^3-\frac{1}{2}u^4+\frac{1}{5}u^5\right).$$

Faulhaber的实际算法并不完全如此，因为我们用到的$u=2\sum n$,而不是$\sum u$。但是,这点差别当然是微不足道的。

参 考 文 献

[1] A.W.F.Edwards, Sums of powers of integers, *Math. Gaz.*, 66(1982), 22—28.

[2] J.Faulhaber, Academia Algebrae, Augspurg, 1631.

[3] C.G.J.Jacobi, De usu legitimo formulae summatoriae Maclaurinianae, *J.Reine Angew. Math.*, 12(1834),263—272.

[4] I.Schneider,Potenzsummenformeln im 17 Jahrhundert, *Historia Math.*, 10(1983), 286—296.

[5] L.Tits, Sur la sommation des puissances numériques, *Mathesis*, 37 (1923), 353—355.

(朱学贤编译，潘承彪校)

算术平均值-几何平均值不等式的再讨论①

D.P.Minassian

最近，在给学习保险统计学的学生讲授利息理论时，要用到熟知的以下事实：

命题　n个非负数的算术平均值大于或等于它们的几何平均值，而且等号成立当且仅当这 n 个数都相等。

见文献[1]中的第 39 页。

我给出了命题的一个适合大学一年级学生（他们只学过一元微积分）的证明。这个证明看来是新的，至少知道它的人并不多。而已有文献中包含的证明，或者根本不用微积分，或者用到了较深的微积分知识。

我的证明方法还能被用于得到一些推论，这些推论是新的，至少我在文献中没有见到过。这些推论又促使我去考虑：如果将非负假定减弱，不等式又会怎么样？到目前为止，还很少有人注意到这个问题。

命题的证明　设 $x_1, x_2, \cdots, x_n$ 是 n 个任意的非负数，要证明

① The arithmetic-geometric mean inequality revisited, elementary calculus and negative numbers, *Amer.Math.Monthly*, 94(1987), 977—978.

$$\frac{x_1+x_2+\cdots+x_n}{n}\geqslant\sqrt[n]{x_1x_2\cdots x_n}\,, \tag{1}$$

其中等号成立当且仅当 $x_1=x_2=\cdots=x_n$。

如果其中有一个 $x_i=0$，则不等式(1)显然成立，因此假定每一个 $x_i>0$。

令 $k\equiv x_1+\cdots+x_{n-1}>0$，$c\equiv x_1\cdots x_{n-1}>0$，则 (1) 式等价于去证明

$$f(x_n)\equiv(x_n+k)^n-n^n c x_n\geqslant 0. \tag{2}$$

对 n 用数学归纳法。设

$$k\geqslant(n-1)c^{\frac{1}{n-1}}, \tag{3}$$

其中等号成立当且仅当 $x_1=\cdots=x_{n-1}$。当 $n=2$ 及 $n=3$ 时(3)式显然成立。在(2)式中将 x_n 换成实变量 x 且不考虑它的正负。计算 f 的一阶和二阶导数，发现 $f(x)$仅在

$$x_{\min}=nc^{1/(n-1)}-k \tag{4}$$

处取极小值，且

$$f(x_{\min})=n^n c[k-(n-1)c^{1/(n-1)}]. \tag{5}$$

由归纳假定(3)得 $f(x_{\min})\geqslant 0$。如果 n 是偶数，则 $x_{\min}$ 是 $f(x)$的唯一驻点，从而 $f(x_{\min})$是最小值；但当 n 是奇数时，$f(x)$在某个 $x_{\max}<0$ 处有唯一的极大值且 $x_{\max}<x_{\min}$。但无论何种情形，(2)式对非负的 x_n 必定成立。另外，$f(x_{\min})=0$ 当且仅当(3)式是一个等式，由归纳假定，又当且仅当 $x_1=\cdots=x_{n-1}$。于是，由(4)得到 $x_{\min}$(即 x_n) $=x_{n-1}$。

命题证毕。

下面列出一些推论，它们的证明将作为习题留给读者。公式(2)及其证明实际上说明，负的 x（即 x_n）能使(1)式依然成立。

推论1 固定 n 个数 $\{x_1, x_2, \cdots, x_n\}$ 中的任意 $n-1$ 个且假定它们都是正数（不失一般性，设 x_n 是剩下的数）。设 f 如(2)。则，

(i) 如果 n 是奇数，

(a) 如果至少有 2 个 $x_i\,(i<n)$ 是不相等的，则 f 有唯一的零点 $r<0$；另外，不等式(1)成立当且仅当 $x\equiv x_n>r$，且仅在 $x\equiv x_n=r$ 处(1)是一个等式。

(b) 如果所有的 $x_i\,(i<n)$ 都相等，则 f 恰有 2 个零点：$r_1<0$ 及 $r_2>0$；另外，(1)式成立当且仅当 $x\equiv x_n\geqslant r_1$，且仅在 r_1 和 r_2 处等式成立。

(ii) 如果 n 是偶数，则 f 非负。

这一推论产生了算术平均值-几何平均值不等式的一个未被注意到的推广。

推论1′ 设 $x_1, \cdots, x_n$ 是 n 个非零数（只是为了简便），它们不一定全是正的，其中的 $n-1$ 个数，设为 $x_1, \cdots, x_{n-1}$，有满足(3)式的正和及正积。则

(a) 若 n 是奇数，则对推论 1 中的所有 $x\equiv x_n\geqslant r$（或 r_1），算术平均值-几何平均值不等式成立。

(b) 若 n 是偶数，则算术平均值-几何平均值不等式成立。而且，如果在不等式(1)的两边同时取 n 次幂，则无论 x_n 是正还是负，这种形式的算术平均值和几何平均值不等式总成立。

也许是因为这一结果，鼓舞了我们减弱非负假设去进一步研究算术平均值-几何平均值不等式。我们仍将下面的结论作为习题留给读者，它们是新的结果。

令

$$g(x_1, \cdots, x_n) \equiv (x_1 + x_2 + \cdots + x_n)^n - n^n x_1 x_2 \cdots x_n \qquad (6)$$

是 n 元函数。

推论2 设 n 是奇数，$x_1, \cdots, x_n$ 是任意实数。又设除了第一卦限及坐标均为负数的卦限外，在 R^n 的其余 $2^n - 2$ 个卦限的任意一个中，g 都在 $(-\infty, \infty)$ 中变化。在(2)式中如果 $c = 0$ 且 x_i $(i < n)$ 中恰好奇数个非正或者其中有一个是零，则方程 $g = 0$ 隐含 x_n 是其余变量 $x_1, \cdots, x_{n-1}$ 的函数，即 $g = 0$ 是"唯一的" $n - 1$ 维超曲面。

推论3 设 n 是偶数。则在 n 维 Descartes 空间中的每一条平行于坐标轴的直线上，可能除去一个有限区间外，不等式(1)的两边取 n 次幂后仍成立。另外，除了第一卦限及坐标全为负数的卦限外，g 在其余的每一个卦限上都是无界的。

参 考 文 献

[1] S.G.Kellison, The theory of interest, Irwin,Homewood, IL, 1970.

（朱学贤编译，潘承彪校）

因子分解与素数判定(二)

J.D.Dixon

§7 因 子 分 解

在实践中，当我们要对 n 作素因子分解时，常先用试除法来去掉它的（约不超过 10^4 的）小素因子。然后对未分解的部分反复用伪素数检验法(47)式，直到证实它是合数或确信它必为素数时为止。在确信 n 是素数时，必须给出一个严格的证明，我们将在§13中加以叙述。在证实 n 为合数时，必须要找出 n 的一个真因子。在找到 n 的一个真因子后，原来的问题就化为两个较简单的问题，然后重复上面的做法。迄今所知分解因子的最强有力的方法可在数小时内将50位左右的数作因子分解，但是对再大的数进行分解就要碰碰运气了(见[45])。对大数成功地加以分解是用一组特定的方法来实现的，其中某种方法可能对某类整数或多或少更适合一些。在§8—11中将对实践中有一定实效的主要方法概略予以介绍，其细节可在[20]及[21]中找到。在§12将讨论一些理论性的结果。

§8 直接法和筛

一旦从给定的整数 n 中去掉了它的小素因子，试除法通常就不再有效了。筛法是一种改进，它可以剔除各种类型的

可能因子。比方设 n 是一个 Fermat 数 $F_k=2^{2^k}+1$，有素数 $r|n$。那么 U_r 中的元素 $x=2\bmod r$ 满足关系式 $x^{2^k}=-1$，由引理 1 知，x 的阶为 2^{k+1}。因此 $2^{k+1}|(r-1)$。这也就是说，n 的所有素因子 r 都在算术级数 $2^{k+1}m+1(m=1,2,\cdots)$ 中。

习题13 (a) 证明，若 $k\geqslant 2$ 且有素数 $r|F_k$，那么就有 $2^{k+1}|(r-1)$。$\left[\text{提示：} r\equiv 1(\bmod 8)，\text{故}\left(\frac{2}{r}\right)=1.\right]$

(b) (Euler,1747) 求 F_5 的一个素因子。

显然，刚才所说的方法与 n 的特殊形式有很大关系。一种更广泛适用的筛法要用到二次剩余。1798 年 Legendre 发现，如果对某两个与 n 互素的整数 a,c 有 $c^2\equiv a(\bmod n)$，则对每个整除 n 的素数 r 皆有 $c^2\equiv a(\bmod r)$，从而 $\left(\frac{a}{r}\right)=1$。如果 a 为奇数，则由二次互反率知 $\left(\frac{r}{|a|}\right)$ 可以算出来，这就说明 r 必定在某个算术级数中，当 $4|(a-1)$ 时这级数的公差为 $|a|$，而当 $4\nmid(a-1)$ 时，这级数的公差为 $4|a|$。如果知道模 n 的若干小平方数，则会对界限 r 的可能取值有所帮助（见[23]。类似的方法出现在[17] § 322中）。

例 作为一个稍显造作的例子，我们考虑 $n=1711$。注意 $16^2\cdot 7-9^2=1711$ 及 $37^2\cdot 5-1=4\cdot 1711$，故 7 和 5 均为模 n 之平方，故对每个整除 n 的素数 r 有 $\left(\frac{r}{5}\right)=\left(\frac{5}{r}\right)=1$ 以及

$$\left(\frac{r}{7}\right)=\left(\frac{7}{r}\right)(-1)^{(r-1)/2}=(-1)^{(r-1)/2}.$$

从而 $r\equiv\pm 1(\bmod 5)$ 且 $r\equiv\pm 1,\pm 9$ 或 $\pm 25(\bmod 28)$。同时

满足这两组条件的最小的一些整数是 1,9,19,29,…，实际上确有 $r=29$ 整除 n。

此例表明，在应用这个方法时会出现两个困难。首先，要能找到一些比较小的整数，它们是模 n 的平方数，第二，需要有一种有效的方法，能将对不同的模所获得的有关 r 的信息综合起来。Legendre 建议对各个整数 k 用 $(kn)^{1/2}$ 的连分数逼近来解第一个问题（见下面第 11 节），在过去的五十年里，Lehmer 及其同事造出一系列特殊用途的计算机（“延迟线筛”），用以解决第二个问题。渐近地说来，此法很差，其运算时间的增长性大致与 $cn^{1/2}$ 相当，不过这里常数 c 可能相当小。尽管如此，这个方法一直颇具竞争力，因为对完成它们的特殊任务来说，延迟线筛要比当代的通用计算机要快得多。七十年代初期在 Berkeley 建造的 SRS-181 每秒可处理 $2\cdot10^7$ 个整数，此后人们又造出一些更快的筛（例如 Williams 在 Winnipeg 所造者）。

§9 Pollard 的 Monte Carlo 方法（或 Rho 方法）

设 S 是一个有限集合，Map(S)是由 S 到 S 的所有映射组成的集合，对任两个从 S 到 S 的映射 $\lambda,\mu\in$ Map(S)，定义 λ 与 μ 的一个复合运算 $\circ$ 如下：

$$(\lambda\circ\mu)(x)=\lambda(\mu(x)),\quad \forall x\in S.$$

任取一个映射 $\psi\in$ Map(S) 及一个元 $x_0\in S$，按

$$x_0=\psi^0(x_0),\psi(x_0),\psi^2(x_0)=\psi(\psi(x_0)),\cdots$$

一直做下去，就得到属于 S 的一列元素，但由于 S 中仅有有限个元素，故必存在一个整数 $l\geqslant 1$，使 $x_0,\psi(x_0),\cdots,\psi^{l-1}(x_0)$ 皆不相同，但有某个 $l'\in[0,l-1]$ 使 $\psi^l(x_0)=\psi^{l'}(x_0)$。

数 l 称为 x_0 在映射 ψ 下的轨道长度。注意，当 $k \geqslant l'$ 时，序列$\{\psi^k(x_0)\}$是以 $l-l'$ 为周期的。

现在设 S 有 m个元素且 $x_0 \in S$。只要 $x_0, \psi(x_0), \cdots, \psi^{l-1}(x_0)$皆不相同，$x_0$ 在映射 ψ 下的轨道长度就至少是 l。故使 x_0 的轨道长至少为 l 的那种映射 $\psi \in \mathrm{Map}(S)$的个数为

$$m^{m-l}\prod_{i=0}^{l-1}(m-i)<m^m\exp\left(-\sum_{i=0}^{l-1} i/m\right)$$

$$=m^m\exp(-l(l-1)/2m).$$

因此，对每个 $\lambda>0$，$\mathrm{Map}(S)$中使 x_0 的轨道长至少为$(2\lambda m)^{1/2}+1$ 的映射 ψ 所占比例少于 $\exp(-\lambda)$。于是，将只有比较少的元素对 x_0,ψ 使其轨道长大于比方说 $5m^{1/2}$，我们将把这种元素对 x_0,ψ 称为“例外的”(与[21]习题 3.1.12 比较)。

应用这一事实，J.M.Pollard 在[39]及[40]中得到如下的因子分解方法。设 n 为合数且 $r|n$。令 $S=\boldsymbol{Z}_r$，设 y_0 为一个整数，ψ 是一个多项式函数。如果元素对 $y_0(\bmod r),\psi$ 不是例外的，$y_0(\bmod r)$在映射 ψ 下的轨道长度不超过 $r^{1/2}$ 的一个小倍数。我们不知道 r，因而无法计算 $\boldsymbol{Z}_r$ 中的诸元素 $x_k=\psi^k(y_0 \bmod r)$，但我们可以计算序列$\{y_k\}$，其中对每个 $k>0$，$y_k \in [0,n-1]$且 $y_k \equiv \psi(y_{k-1})(\bmod n)$；又对所有 k 皆有 $x_k \equiv y_k(\bmod r)$。利用上述记号即知，只要 $k \geqslant l'$，$\{\psi^k(x_0)\}$就是以 $d=l-l'$ 为周期的序列。这表明，只要 $k>h>l'$ 且 $d|(k-h)$，就有 r 整除 $\mathrm{GCD}(y_k-y_h,n)$。这是求递推序列中的圈的一种特例。首先是 R.W.Floyd 求解这一问题，他发现可以对 $k=2h$ 求出解答，此后，Brent[6]证明了，对形如 $h=2^i-1, k=2^i+j, 0 \leqslant j<2^i$ 的 k 及 h 求解常常更好些。

其关键在于按此法求合适的 y_h 及 y_k 只需对该序列中的两个值保留记忆，故而 Pollard 的方法只要求很少的存储量。除非我们极不走运，否则总可以得到 n 的一个形如 $\mathrm{GCD}(y_k-y_h,n)$ 的真因子。

例 设 $n=30623$，取 $\psi(x)=x^2+1$ 及 $y_0=1$。那么，对所有 $k>0$，满足 $y_k\equiv\psi(y_{k-1})(\bmod n)$ 的 y_k 的值列即为：1,2,5,26,677,29608,19667,22400,…。实际上，当 k 取遍区间 $[2^i,2^{i+1}-1]$ 时，只有 y_k 的现在数值和 y_{2^i-1} 的值储存着。我们发现，$y_7-y_3=22374$ 与 n 有一公因子 113，故 113 即为 n 的一个真因子。

怎样来挑选 ψ 和 x_0，使之作成对 Z_r 的非例外元素对呢？尽管可以证明 ψ 应是非线性的，也不能形如 x^2-2，但一般来说对此没有什么肯定的结论。对直到 10^6 的素数提供的数据（见[20]）表明，$\psi(x)=x^2+1$ 或 x^2-1 既容易计算，也相当好，这些函数用得最广，通常在 $O(r^{1/2})$ 步后即会得出素因子。此法最令人注目的成功是对 78 位 Fermat 数 F_8 作因子分解，F_8 是合数，这已于 1909 年被 J.C.Morehead 和 A.E.Western 所证明(见[8])。Pollard 证明了 F_8 有下列分解式：

$$F_8=934616397153579777691635581996068965840512375416381885802803 21\times 1238926361552897.$$

有关 Pollard 方法的细节及实施等，见[6],[8],[20],[21]及[40]。

既然每个合数 n 皆有一个素因子 $r\leqslant n^{1/2}$，Monte Carlo 方法的执行时间大约为 $O(n^{1/4})$，但是这个结论是以一个未能证明的猜想为条件而获得的，这个猜想是说所选取的特殊元

素对 x_0,ψ 是非例外的。至于这个方法（或者是它的某种修改）是否可进行完全的分析，这仍是一个没有解决的问题。

习题14 证明，若 $x_0\in S$ 且 $|S|=m,\psi$ 取遍从 S 到 S 的所有一对一的映射，则在这种映射 ψ 下，x_0 的平均轨道长度为 $(m+1)/2$（此例可用来解释为什么 Pollard 方法中不应用线性多项式）。

习题15 （Pollard） 设 p 为素数，g 是模 p 的一个原根。利用上面的思想给出一个算法，使对任一满足 $p\nmid a$ 的整数 a，此法可求出关系式 $g^i\equiv a(\bmod p)$ 中的指数，其执行时间以 $O(p^{1/2}(\ln p)^3)$ 为界（有关结果见[3]及[38]）。

§10 利用 U_n 的群构造求 n 的因子

若 n 有 s 个不同的素因子，由于对每个 $n_i=r_i^{k_i}$，U_{n_i} 都有一个唯一的 2 阶子群，因而不难证明 $T=\{x\in U_n|x^2=1\}$ 是 U_n 的一个阶为 2^s 的子群（见(49)式）。反之，若对某个整数 c 有 $c^2\equiv1(\bmod n)$ 且 $c\not\equiv\pm1(\bmod n)$，则 $\mathrm{GCD}(c-1,n)$ 是 n 的一个真因子。因此 T 中每个元素 $x\neq\pm1$ 都产生出 n 的一个真因子，许多因子分解方法都是以此作为基础的。

习题16 设对某个整数 $d>1$，若 $\boldsymbol{Z}_n[X]$ 中一个 d 次多项式在 $\boldsymbol{Z}_n$ 中有 $d+1$ 个不同的根，那么，你可以怎样来求 n 的一个真因子呢？

现在假设知道有一个整数 $m>0$，它对所有 $x\in U_n$ 满足 $x^m=1$。记 $m=2^hl,2\nmid l$。如果 n 不是一个素数幂，则引理4说明：对 U_n 中至少一半元素 x，循环群 $\langle x\rangle$ 是偶数阶的，但并不包含 -1。对于这种元素 $x\in U_n$，存在一个指数 $t\in[0,h-1]$，使 $y=x^{l2^t}\neq\pm1$，但 $y^2=1$，因此可以得到 n 的一个

真因子。这就产生出一种概率算法，在这个算法中，从U_n中独立并一致地随机选取一列x的值，当找到一个x，能按上法从x得到n的一个真因子，则算法中止。需要选取的x的个数平均来说至多为2，故而为求得n的一个因子所需的单精度运算的平均次数是$O((\ln n)^2 \ln m)$。例如，如果知道U_n的阶$\varphi(n)$，则$m=\varphi(n)$可很快对n加以分解。取$m=n!$也行得通，不过那样的话$\ln m$就太大了。

一般说来，不可能找到一个足够小的m使上述方法可行，除非能对n的算术构造有进一步的了解（而这正是我们要寻求的！）。尽管如此，D.N.Lehmer和D.H. Lehmer曾提出过此法的一个变体（未发表），J.M.Pollard[39]也提出过此法的变体，这些方法都获致某种成功。固定一个常数$c>0$（在某些应用中大约取为10^5），令m表示小于c的所有素数幂的最小公倍数.若n有分解式(1)且对所有i有$\varphi(r_i^{k_i})|m$，则由(49)式知，对所有$x\in U_n$皆有$x^m=1$，故上面的方法适用。然而，如果有至少一个$\varphi(r_j)|m$，我们仍能得到一个分解，因为在此情形我们总有$x^m \bmod r_j=1$，因而$GCD(x^m-1,n)\neq 1$.进一步的修改(见[20],[39]及[63])扩大了这个方法的适用范围，但是方法能否成功在一定程度上依赖n的取值。

例　设$n=1711$并取$c=10$，则$m=2^3\cdot 3^2\cdot 5\cdot 7=2520$。取$x=2$，则$x^m\equiv 523 \pmod n$且$GCD(522,n)=29$，这就是$n$的一个真因子。

习题17　如果k与U_n的阶互素，证明$x\mapsto x^k$是从U_n到其自身的一个一对一的映射，其逆映射也有$x\mapsto x^{k'}$的形式，这里k'是某个整数。试说明，如果n的素因子已知，

怎样从 k 来计算 k'. 反过来证明，如果已经知道一对这样的整数 $k,k',k>1$（它们的大小与 n 相当），n 可以很容易加以分解。(因此计算反函数与将 n 分解成素数一样困难.这些映射被用作"陷门函数"，例如见[13]。)

§11 因子基和连分数方法

如上一节所述，只要能找到 U_n 中一个满足 $x^2=1$且$x\neq\pm1$的元 x，就能求得 n 的一个真因子；反过来，只要 n 不是素数幂，就存在这样一个元素 x。显然，只要求满足 $a^2\equiv b^2 \pmod n), a\not\equiv\pm b \pmod n$的整数 a,b 就行了，因为GCD$(a-b,n)$还是 n 的一个真因子。1643年，通过直接求满足 $n=a^2-b^2$ 的 a 和 b，Fermat 对整数进行了因子分解（见[12]p. 357)。最近 R.S.Lehman[25]证明了，对 Fermat 的想法用 Farey 分数加以修改即可给出一种分解因子算法，其运算时间为 $O(n^{1/3})$，但它似乎还不能与其它已知的算法匹敌。一种更好的方法如下所述。

我们定义因子基是由一组非零整数组成的一个集合 $B=\{b_0,b_1,\cdots,b_h\}$，例如由 -1及前 h 个素数组成的集合。整数a称为一个 B-数，如果由 $c\equiv a^2 \pmod n$及$c\in\left[-\frac{n}{2},\frac{n}{2}\right]$所定义的整数 c 可表为B中若干因子的乘积。故若 a 是一个 B-数，我们就可算出整数 $e(a,i)\geqslant0$使 $a^2\equiv\prod_{i=0}^{h} b_i^{e(a,i)} \pmod n$。对每个 B-数 a，相应定义 $h+1$ 维向量 $e(a)=(e(a,0),\cdots,e(a,h))$。如果能求得由 B-数组成的一个充分大的集合A（只要A的元素个数$|A|=h+2$就足够了），那么对模 2 来说，**诸向量 $e(a)(a\in A)$就是线性相关的。应用(Z_2上）通常**

的线性代数知识，可以求得非空子集 $A'\subset A$ 使

$$\sum_{a\in A'} e(a)\equiv(0,\cdots,0)\pmod 2,$$

从而对每个 i，$h_i=\frac{1}{2}\sum_{a\in A'} e(a,i)$ 是整数。现在如果我们定义

$$u\equiv\prod_{a\in A'} a\pmod n,\quad v\equiv\prod_{i=0}^{h} b_i^{h_i}\pmod n,$$

则有 $u^2\equiv v^2\pmod n$。如果碰巧有 $u\not\equiv\pm v\pmod n$，我们就得到 n 的一个真因子。

例 令 $n=1711$，则有

$$41^2-1711=-30,\quad 83^2-4\cdot1711=45,$$

$$124^2-9\cdot1711=-23,\quad 455^2-121\cdot1711=-6.$$

利用基 $B=\{-2,3,5\}$ 即得

$$41^2\equiv(-2)\cdot3\cdot5,\ 83^2\equiv3^2\cdot5,\ 455^2\equiv(-2)\cdot3\pmod n.$$

虽然这里只有三个同余式，但是我们发现上面三式右方由 -2,3及5的指数所组成的三个向量(1,1,1),(0,2,1)及(1,1,0)对模2来说是线性相关的，这是因为这三个向量的和为 $2(1,2,1)\equiv(0,0,0)\pmod 2$。故

$$(41\cdot83\cdot455)^2\equiv((-2)\cdot3^2\cdot5)^2\pmod n,$$

但由于

$$41\cdot83\cdot455\equiv-(-2)\cdot3^2\cdot5\pmod n,$$

我们很不走运。然而，利用另一个等式 $185^2-20\cdot1711=5$ 得到又一个指数向量(0,0,1)。由于

$$(0,2,1)+(0,0,1)=2(0,1,1),$$

我们即得$(83\cdot185)^2\equiv(3\cdot5)^2(\bmod n)$，由此就得出一个真因子为$GCD(83\cdot185-3\cdot5,n)=29$。

这个想法可追溯到1926年 M.Kraitchik 的工作[22]，不过它并未以任何系统的方式表述过。1931 年，D.H.Lehmer 和 R.E.Powers[31]提出一个类似的想法，其中要用到 Legendre 的一个方法来产生比较小的二次剩余（可以预料它们常可分解成较小素数的乘积）。但由于当时还没有快速计算工具，它们的方法无法实用。直到出现储存量大的快速电子计算机，情况才有了改变。六十年代后期，以[31]的思想作为基础，J.Brillhart 和 M.A.Morrison 将其系统化并在计算机上实际使用。当即获得首次重大成功：将39位的Fermat数 F_7 进行了因子分解，他得到

$$F_7=(59649588127497 21)\cdot(5704689200685129054721).$$

自1897年以来（见[36]）人们就已知道 F_7 是合数。这种方法现在称为连分数方法，在对直到大约50位的整数进行因子分解时，它是当今最成功的一种方法。

那么，Legendre 求小二次剩余的思想是什么呢？一般来说，任何无理数的连分式展开都给出一列越来越接近的有理逼近（比方见[61]第 1 章）。特别地，如果 n 不是完全平方数，只利用整数的算术，可以算出两个递增的整数序列$\{u_k\}$和$\{v_k\}$，使对每个 k 有

$$|n^{1/2}-u_k/v_k|<1/v_kv_{k+1},$$

故有 $u_k^2-nv_k^2=w_k$，这里 w_k 是一个整数且$|w_k|<2n^{1/2}$。因此$\{w_k\}$是 n 的一列二次剩余，其大小均为 $O(n^{1/2})$。此外，对任何能整除 w_k 的素数 p，n 都是 p 的二次剩余。取因子基

B是由-1及前h个满足$\left(\frac{n}{p}\right)=1$的素数组成的集合（$h$是一个适当选取的正整数），$w_k$比较小时就增加了$u_k$是$B$-数的可能性，这样可望产生出足够多的$B$-数以使我们的方法能获致成功。在实践中要注意几个要点。不需要计算出v_k，并且只需对模n计算u_k；有时候对某个小整数d用$(dn)^{1/2}$的连分式来代替$n^{1/2}$会有所裨益。对此人们做过许多研究，以设法保证算法的实施尽可能有效，其中一些成果见原始论文[36]。文[64]中的数据显示，当n在10^{30}到10^{50}这范围内时，平均运算时间大约与$n^{1/7}$成比例。最近的论文[45]讨论了若干新方法，它们使这算法的运算速度提高到原来的10—15倍，其中一种方法是"早期夭折战略"，即当有证据显示w_k没有足够小的因子时，就在早期弃去w_k。迄今尚无人对连分数算法的有效性进行过完全的理论分析，但是最近的两篇文章证明了：以一些合理的但未经证明的关于素数分布的猜想为基础，该算法的渐近运算时间形如

$$\exp\{c(\ln n\ \ln\ln n)^{1/2}\}=n^{C(\ln\ln n/\ln n)^{1/2}}$$

（见[43]及[52]）。

§12　因子分解问题的理论估计

令人遗憾的是，实践中最有效的那些因子分解方法的性能均无恰当的分析。如上所述，连分数方法颇得经验证据的支持（见[64]与[45]）。[63]对第10节中讨论的方法给出某些统计资料。[20]中报告了在极限范围内支持Monte Carlo方法的计算证据。在有关素数和素因子分布的某些合理的但是未经证明的猜想成立的条件下，[43]和[52]都对连分数方

法及其变种进行了分析。但这仍然不能保证对 n 的某些（也可能是大多数）值，这些方法不会全都变得很差。于是人们很希望能有一种（虽然从实践的观点看它肯定会是很差的）方法，对这方法可以作完全的分析，并能证明对每个 $\varepsilon>0$ 其运算时间是 $O(n^{\varepsilon})$。

这种渐近快速因子分解算法是一种概率算法，它是以自然利用因子基为其基础的。利用第11节的记号，我们考虑由 -1 和前 h 个素数组成的基 $B=\{b_0,\cdots,b_h\}$。这个算法中的基本步骤在于从 $[1,n-1]$ 中独立并一致地随机选取足够多的整数，直到得到由 B-数组成的一个集合 A，其指数向量 $e(a)(a\in A)$ 对模 2 是线性相关的。如第11节所指出的，这样就产生出一对满足 $u^2\equiv v^2(\bmod n)$ 的整数 u,v，当 $u\not\equiv\pm v(\bmod n)$ 时就得到 n 的一个真因子。现在证明下面的定理就不太困难了。

定理7 （J.D.Dixon） 设 n 不是素数幂，则上面所述之基本步骤能找到 n 的一个真因子的概率至少为1/2。此外，存在绝对常数 $c_1,c_2>0$，使得如果 $h=[\exp\{c_1(\ln n\ln\ln n)^{1/2}\}]$ 且上述基本步骤一直重复到求得 n 的一个真因子为止，则算法的平均运算时间为 $O(\exp\{c_2(\ln n\ln\ln n)^{1/2}\})$。

注 在原文[14]中得到 $c_1=\sqrt{2}$，$c_2=3\sqrt{2}$。这些数值在[21]与[52]中得到了改进。在[43]中证明了，对于最佳选择的 h，平均运算时间形如 $\exp\{(2+\varepsilon_n)(\ln n\ln\ln n)^{1/2}\}$，这里 $\varepsilon_n\to0(n\to\infty)$.因此运算时间的增长性比 $\ln n$ 的指数函数慢，而比 $\ln n$ 的任何幂次增长得都快。

文[43]与[52]对整整一类因子分解算法进行了仔细的分析，它们皆属上面的类型，但用了不同的方法以使能更有效

地选取由 B-数组成的集合 A。这些论文使得用来支持所提出的方法的某些非常粗略的讨论得以精确化。对算法的分析是不完全的（这些分析是以清楚陈述的但未经证明的合理猜想为基础的），但它们的价值在于显示出各种不同的修改可能会怎样影响算法的实施。这两篇论文在处理连分数方法的同时，还都对这里未曾讨论过的一些因子分解方法作了颇有趣味的详细研究，其中包括 R.Schroeppel 的线性筛（见[43]），Pomerance 的二次筛（见[43]及[18]），以及 D.Shanks 的一种方法的变形（见[56]，[52]及[53]），Shanks 的这一方法在判别式为 $-n$ 的二元二次型的等价类作成的群中有效（这把我们带回到 Gauss 最初的研究[17]）。我们还要提及Shanks 的另一个称为 SQUFOF 的方法，它对某种范围的 n 相当有效（见[59]）。

§13 寻求近似多项式时间的素性证明

分解整数的最后一步是证明分解出的各个因子均为素数。假设 n 是猜想为素数的一个大整数。我们需要对它作足够多的(47)型之伪素数检验，直到确信 n 极可能是素数时为止，而这时我们面临的任务是要找到一个证明。直到不久以前，虽然常可以对80位或更多位数的某种素数给出素性证明，但却并没有比较快且一致的方法，常常较小的素数也会弄得我们束手无策（见[62]）。然而在 1980 年，Adleman、Rumely 提出了证明素性的一个普遍有效的算法，利用 K. Prachar 及 P.X.Gallagher的研究结果，A.Odlyzko 与 Pomerance证明了：该算法运算时间接近多项式时间，即为 $O((\ln n)^{c\ln\ln\ln n})$，这里 $c>0$ 为一个适当的常数（见[4]）。

H.W.Lenstra,Jr.和H.Cohen作出此算法的一种流线形式，而Cohen，A.K. Lenstra及D. T. Winter对它仔细地编制了程序（见[11]及[33]）。最新的报告说，此程序可用不到十分钟的时间对直到200位的素数循常规给出证明。尽管如此，是否存在一种可用严格多项式时间给出素性证明的算法，仍是一个没有解决的问题。

Adleman-Rumely算法的基本思想可以叙述如下。我们需要在分圆域中加以讨论，本处所用的方法虽已比较初等，但仍希望读者能对分圆域及Gauss和的有关理论有一定的了解。现在想要对n施行一系列推广的伪素数检验法。如果其中有任何一个检验未能通过，n就不是素数；如果能通过所有这些检验，我们就利用这些检验所得之信息来构造一个筛，用这个筛来对n的可能的素因子予以限制（与第8节比较之）。如果n是一个素数，那么，由于对这种情形中我们所处理的环的算术结构我们知道得很多，因而可以选择检验法以做出一个非常简单的筛。如果这些检验法令人满意的话，我们就能证明n的每个素因子r必定位于比较小的一组算术级数中，这些算术级数形如$t^i \pmod w$，$i=0,1,\cdots,w>n^{1/2}$。此外，选择w可以保证U_w的阶至多为

$$z=O((\ln n)^{c\ln\ln\ln n}),$$

故而只需对i的小于z的值加以考虑。于是，为了证明n是素数（或者最终发现n不是素数），只需验证n在上面所述那组数中没有真因子。（在下面考虑的概率形式中可取$t=n$。）此算法的关键部分是怎样把用伪素数检验法所得来的那么许多信息转移到筛上？一开始（见[4]）这是通过一个一般性的互反律来实现的 Eisenstein 互反律就够用了），但

在下面给出的形式里我们按[33]的做法，不用互反律而改用基于 Gauss 和的初等方法。

下面的定理8描述了我们所谓的广义伪素数检验法。为了说明这些检验法并给定理的证明作准备，我们首先讨论 Gauss 和的某些初等性质（见[20a]）。

设 p 和 q 是不整除 n 的两个素数，$p|(q-1)$，设 ζ_p 与 ζ_q 为 $\boldsymbol{C}$ 中的 p 次及 q 次本原单位根。设 $g=g_q$ 是 $q-1$ 阶循环群 U_q 的一个生成元。由于 $p|(q-1)$，定义 χ_{pq} 是从 U_q 中元素到循环群 $\langle\zeta_p\rangle$ 中元素的一个映射：$\chi_{pq}(g^j)=\zeta_p^j$（对所有 j）。容易验证 χ_{pq} 是 U_q 到 $\langle\zeta_p\rangle$ 的一个同态，它称为以 q 为前导子（conductor）的 p 次特征。对每个整数 t 我们定义 Gauss 和

$$\tau(\chi_{pq}^t)=\sum_u \chi_{pq}(u)^t\zeta_q^u,$$

其中 u 取遍 U_q 中所有元素（表达式 ζ_q^u 有明显的意义，而 ζ_q^j 的值只与 $j \bmod q$ 的值有关，故无任何含混不清之处）。注意 $\tau(\chi_{pq}^t)$ 在环 $R_{pq}=\boldsymbol{Z}[\zeta_p,\zeta_q]$ 中。

引理6 将 χ_{pq} 写成 χ，我们有

(a) $\tau(\chi)\tau(\chi^{-1})=\chi(-1)q$；

(b) 若 n 为素数，则在 R_{pq} 中有

$$\tau(\chi)^{n^{p-1}-1}\equiv\chi(n)\pmod n;$$

(c) 若对每个素数 $r\neq p$ 有 $\zeta_p^i\equiv\zeta_p^j \pmod r$，则必有

$$\zeta_p^i=\zeta_p^j.$$

证明 以下出现的所有的和式中，求和变量皆遍取 U_q 中所有元素。我们要多次利用如下事实：如果 v 是 U_q 中一个固定的元素，那么当 u 遍取 U_q 中一切元素时，uv 也遍取

U_q中一切元素。这等价于说：设 v 是一个与 q 互素的整数，那么当 u 遍取模 q 的一个简化剩余系时，uv 也恰取遍模 q 的一个简化剩余系。简记这结果为(F)。

(a) 的证明：我们有

$$\tau(\chi)\tau(\chi^{-1}) = \sum_u \sum_v \chi(u)\chi(v)^{-1}\zeta_q^{u+v}$$

$$= \sum_u \sum_v \chi(uv)\chi(v^{-1})\zeta_q^{uv+v} \quad (\text{用到}(F))$$

$$= \sum_u \chi(u) \sum_v \chi(1)\zeta_q^{(u+1)v}.$$

由于 $\zeta_q^q = 1$，按等比数列求和易有

$$\sum_v \chi(1)\zeta_q^{(u+1)v} = \sum_v \zeta_q^{(u+1)v} = -1 + \sum_{v=1}^{q} \zeta_q^{(u+1)v}$$

$$= -1 + \begin{cases} (\zeta_q^{q(u+1)} - 1)/(\zeta_q^{u+1} - 1) = -1, \\ \qquad q \nmid (u+1)\text{时}, \\ q = q-1, \quad q \mid (u+1)\text{时}. \end{cases}$$

注意，当 u 取值在集合$\{1,2,\cdots,q-1\}$中时，使得$q\mid(u+1)$的只有 $u=q-1$ 这一个值，于是

$$\sum_u \chi(u) \sum_v \chi(1)\zeta_q^{(u+1)v} = (q-1)\chi(q-1) - \sum_{\substack{u \\ u \neq q-1}} \chi(u)$$

(注意 χ 以 q 为周期)

$$= (q-1)\chi(-1) - \sum_u \chi(u)$$

$$+ \chi(-1).$$

另一方面，易见对 U_q的生成元 g 有

$$\chi(g)=\zeta_p\neq 1,$$

从而由

$$\sum_u \chi(u)=\sum_u \chi(gu)=\chi(g)\sum_u \chi(u)$$

即得$\sum_u \chi(u)=0$，于是得到

$$\tau(\chi)\tau(\chi^{-1})=q\chi(-1).$$

(b) 的证明：为使记号简化，暂令$m=n^{p-1}$。由于n为素数，对所有$a_1,\cdots,a_t\in R_{pq}$及$i>0$有

$$(a_1+\cdots+a_t)^{n^i}\equiv a_1^{n^i}+\cdots+a_t^{n^i}\ (\bmod n).$$

于是

$$\tau(\chi)^m=\left(\sum_u \chi(u)\zeta_q^u\right)^m\equiv\sum_u \chi(u)^m\zeta_q^{um}\ (\bmod n).$$

由于$p\mid(m-1)$，即有$\chi(u)^m=\chi(u)$。又存在$v\in U_q$使$vm=1$，这是因为$q\nmid m$。于是

$$\sum_u \chi(u)^m\zeta_q^{um}=\sum_u \chi(u)\zeta_q^{um}=\sum_w \chi(vw)\zeta_q^w\ (\text{利用}vm=1)$$

$$=\chi(v)\tau(\chi)=\chi(n)^{1-p}\tau(\chi)\ (\text{由}v=m^{-1}\text{及}m=n^{p-1})$$

$$=\chi(n)\tau(\chi)\ (\text{由定义有}\chi(n)^p=1).$$

最后，由于$\chi(-1)q=\pm q$与n互素，则存在一个整数a使$\chi(-1)qa\equiv 1(\bmod n)$。在同余式

$$\tau(\chi)^m\equiv\chi(n)\tau(\chi)\ (\bmod n)$$

两边用$a\tau(\chi^{-1})$乘之，再利用(a)即得欲证之结论。

(c) 的证明：令$k=i-j$。则(c)中假设条件蕴含$\zeta_p^k\equiv 1(\bmod r)$，故

$$f(\zeta_p^k) \equiv f(1) = p \pmod r,$$

这里 $f(X)$ 是所谓的分圆多项式

$$(X^p - 1)/(X - 1) = X^{p-1} + X^{p-2} + \cdots + 1.$$

由于 $r \nmid p$，故 $f(\zeta_p^k) \neq 0$，从而 ζ_p^k 不是 p 次本原单位根. 因此 $\zeta_p^k = 1$，故 $\zeta_p^i = \zeta_p^j$. 这就完成了引理的证明.

现在设 P 与 Q 是由不能整除 n 的素数组成的有限集，其中对所有 $q \in Q$，$q-1$ 都是 P 中不同素因子的乘积. 置 $z = \prod_{p \in P} p, w = \prod_{q \in Q} q$. 那么，对每个 $q \in Q$ 有 $q-1 \mid z$，因而 U_w 中每个元素的阶必定整除 z（与(49)比较之）. 于是，对每个与 w 互素的整数 a 有 $a^z \equiv 1 \pmod w$.

定理 8 (H.W.Lenstra Jr.) 假设 n 是满足下列条件的一个奇整数：

(i) 对所有素数 $p, q, q \in Q, p \mid (q-1)$，有

$$\tau(\chi_{pq})^{n^{p-1}-1} \equiv \chi_{pq}(n) \pmod n;$$

(ii) 对每个素数 $p \in P$ 及每个素数 $r \mid n$ 有

$$p^{e_p} \mid (r^{p-1} - 1),$$

这里 p^{e_p} 是 p 整除 $n^{p-1}-1$ 的最大幂次.

那么，对每个素数 $r \mid n$，皆有某个 $i \in [0, z-1]$ 使 $r \equiv n^i \pmod w$.

注 若 n 为素数，则引理 6 的 (b) 表明条件 (i) 是满足的，至于 (ii) 是显然满足的. 能满足这些条件的其它整数可能很少.

证明 条件 (ii) 表明，对每个 $p \in P$ 和每个素数 $r \mid n$，都存在整数 a_p 及 $b_p(r)$ 使

$$n^{p-1}=1+a_p p^{e_p}, \qquad r^{p-1}=1+b_p(r)p^{e_p},$$

且 $p\nmid a_p$。从而对某个整数 $l_p(r)$ 有

$$a_p l_p(r)\equiv b_p(r)\pmod p。$$

而对每一对 $p,q,q\in Q,p|(q-1)$，引理 6 (b)表明对每个素数 $r|n$ 有

$$\chi(r)\equiv\tau(\chi)^{r^{p-1}-1}\pmod r,$$

这里再次把 χ_{pq} 写成 χ。然而

$$(r^{p-1}-1)a_p=(n^{p-1}-1)b_p(r),$$

因此，取模为 r 来应用(i)中之同余式即得

$$\chi(r)^{a_p}\equiv\chi(n)^{b_p(r)}\pmod r,$$

故由引理6(c)有 $\chi(r)^{a_p}=\chi(n)^{b_p(r)}$。因为此式的两边均为 p次单位根，且 $p\nmid a_p$，由 $l_p(r)$的取法即得$\chi(r)=\chi(n)^{l_p(r)}$。这对每个素数 $r|n$ 以及所有满足 $p|(q-1)$的$p\in P$ 及$q\in Q$ 均成立。

最后，我们利用中国剩余定理（即孙子定理）来求一个整数 $l(r)\in[0,z-1]$，使对每个 $p\in P$有 $l(r)\equiv -l_p(r)\pmod p$。那么，由于 $\chi(n)$是一个 p 次单位根，我们就有

$$\chi(rn^{l(r)})=\chi(r)\chi(n)^{l(r)}=1,$$

因此 $rn^{l(r)}$ 总是在映射 $\chi=\chi_{pq}$(对所有 $q\in Q$ 及所有 $p|(q-1)$)的核中。由于 $q-1$ 无平方因子，故当 p 跑遍 $q-1$ 的素因子时所有这些核的交集只有单位元这一个元。从而对所有 $q\in Q$有 $rn^{l(r)}\equiv 1\pmod q$，故有 $rn^{l(r)}\equiv 1\pmod w$。于是对 $i=z-l(r)$有 $r\equiv n^i\pmod w$，定理证毕。

习题18 设对某个(不一定在Q中的）素数 q 有 $p|(q-1)$，$\chi_{pq}(n)\neq 1$且定理 8 中条件(i)的同余式成立。证明定理 8 的条件(ii)也对 p 成立。[提示：设 h 是环 R_{pq}/rR_{pq}的单

位数群中元素 $\tau(\chi_{pq}) \bmod r$ 的阶，来证明 $p^{e_p+1} | h$。然后利用引理6(b)来证明 $h | (r^{p-1}-1)p$。]

为了应用定理8，需要能对定理的条件进行验证。一种方法是首先用幂算法验证(i)，然后再利用习题18。利用习题18时需要找一个素数 q 使有 $q \equiv 1 \pmod p$，$\chi_{pq}(n) \neq 1$。这在实际上可以很快办到，这是因为（渐近地来说）满足 $q \equiv 1 \pmod p$ 的所有素数 q 中有 $(p-1)/q$ 的 q 也满足 $\chi_{pq}(n) \neq 1$。然而，证明有一个"好的"小素数 q 存在似乎需要假设 GRH 成立才行。在[11]中给出了另一种方法，这种方法避免了上述问题。

剩下的问题是应该怎样选取合适的集合 P 和 Q？为使基于定理8给出的素性证明有效，我们希望 P 与 Q 要小，以使能很快地对条件(i)与(ii)加以验证，但同时又希望 z 很小、w 很大，以便可能的素因子能迅速查出。选取 $w > n^{1/2}$ 可以保证每个剩余类 $n^i \bmod w$ 含有至多一个可能的素因子 $r \leq n^{1/2}$，那样的话，对 n 这么大小的数进行 $O(z)$ 次多精度运算即可完成最后一步。此外，若 z 与 w 均不超过 n，那么 $|P|$ 与 $|Q|$ 皆为 $O(\ln n)$，这样一来，(i)和(ii)皆可在 $\ln n$ 的多项式时间内得以验证。于是，下述定理（其证明见[4]）表明，选取合适的 P 和 Q 可以给出一个算法，对任何 $c > c_2$，其运算时间为 $O((\ln n)^{c \ln\ln\ln n})$。

定理9(Pomerance-Odlyzko)　设 P, Q, z 及 w 定义如上，Q 是由所有满足 $(q-1) | z$ 的素数 q 组成。那么，有可计算的绝对常数 $c_1, c_2 > 0$，如果 $w > n^{1/2} \geq 10$，则 z 的最小可能的值满足

$$(\ln n)^{c_1 \ln\ln\ln n} < z < (\ln n)^{c_2 \ln\ln\ln n}.$$

注 [4]中的计算表明，如果取 P 是由前 6 个素数组成的集，则 $z=30030$，Q 由21个素数组成，而 $w>(5.10^{89})^{1/2}$。类似地，若 P 由前 13 个素数组成，则 z 大约为 3.10^{14}，Q 由 807 个素数组成且 $w>(10^{11356})^{1/2}$。这些都是在定理 9 的意义下的最佳选择。

现在的做法[11]与上面所述算法有几方面不同。第一，它是用 Jacobi 和（在原始文献[4]中引进的）代替了 Gauss 和。验证定理 8 的条件(i)所需要的计算可以用较小的环 $\boldsymbol{Z}_n[\zeta_p]$ 中所作的计算来代替，且指数 $n^{p-1}-1$ 可减少到大致 n 那么大。第二，定理 8 的一个推广允许 z 和 w 有重素因子，这就对实际运算时间有重要的影响，尽管这对渐近情形的运算时间没有多少作用。结果就得到一个有效的素性证明程序，此程序是以对早先方法作出理论及实践上重大改进的一种算法为基础的。最后我们要提到的是，[4]和[33]也对寻求素性证明给出了完全确定的算法，这些算法与上面所述之形式有同样的渐近运算时间。然而，对实际执行来说，已经知道概率算法是更好的选择。

§14 展望未来

就理论方面而言，仍有像"证明素性或分解整数是否有多项式时间算法存在?"这样的问题。证明素性的多项式算法似很有希望存在，但相反地，可能会证实分解因子是 NP-困难的。有关 NP-困难及 NP-完全问题的重要概念在[16a]中作了讨论。

下面是一些研究方向，它们有可能引出更好的理论或实际算法。

1. 是否存在可以快速计算的"高度合成的"整值函数呢？在一种意义上说，这正是第10节中所述方法所蕴含的思想，在那里 x^m-1 可被每个满足 $(r-1)|m$ 的素数 r 整除。例如，若对 $k\in[1,n]$ 可以快速计算 $k!\bmod n$，那么也就可以快速分解 n（有关这个论题的其它研究及结果可参看[55]）。

2. 可否提高第11节中所述方法的速度？比方说，有没有比连分数法更好的方法来定出模 n 的小二次剩余呢？（更小的二次剩余可以增加更大一部分为 B-数的可能性）是否可选取到比小素数更好的因子基 B？有关第11节中基本方法的其它变种，见[43]。

3. 第13节中的 Adleman-Rumely-Pomerance 算法用到 n 通过某种伪素数检验法这一事实来建立一种特别引人注目的筛。当 n 是素数时，是否可能有类似的想法？特别地说，可否构造出一个筛，以使可能的因子都包含在一个可以简洁描述出来的集合中呢？

4. 若不用 $\boldsymbol{Z}_n$ 及 $\boldsymbol{U}_n$ 而改用其它的环和群，又会怎么样呢？有一个"$p+1$"因子分解法，此法是以环 $\boldsymbol{Z}_n[X]/(X^2-c)$ 中的计算为基础的（见[20]p.73 及[63]，其中的想法是以稍微隐蔽的形式用 Lucas 级数表述的）。Shanks 指出过怎样利用二元二次型的类群（见[52],[54],[56],[58]和[59]）。正如第 13 节所示，多一些自由度或许会很有助益。

5. 能否对当今所用的因子分解方法整理出一套理论基础呢？似乎因子分解的一些概率算法会更容易进行分析。

以下所列资料仅为本文中提到的书及论文。更多的参考文献可在[20],[41],[43],[62]以及 Math.Centrum Tracts中找到，文[43]与[54]即发表在 Math.Centrum Tracts 中。

参 考 文 献

[1] W.Adams and D.Shanks, Strong primality tests that are not sufficient, *Math.Comp.*, 39(1982), 255—300.

[2] L.M.Adleman, Two theorems on random polynomial time, *Proc IEEE Symp.Found.Comp.Sci.*, 19(1978), 75—85.

[3] ——, A subexponential algorithm for the dicrete logarithm problem with applications to cryptography, *Proc. IEEE Symp.Found.Comp.Sci.*, 20(1979), 55—60.

[4] L.M.Adleman, C.Pomerance and R.S.Rumely, On distinguishing prime numbers from composite numbers, *Ann.of Math.*, (2)117 (1983), 173—206.

[5] E.T.Bell, Mathematics, Queen and Servant of Science, Bell, London, 1951.

[6] R.P.Brent, An improved Monte Carlo factorization algarithm, BIT, 20(1980), 176—184.

[7] ——, Succinct proofs of primality for factors of some Fermat numbers, *Math.Comp.*, 38(1982), 253—255.

[8] R.P.Brent and J.M.Pollard, Factorization of the eighth Fermat number, *Math. Comp.*, 36(1981), 627—630.

[9] D.A.Burgess, On character sums and primitive roots, *Proc.London Math.Soc.*, (3)12(1962), 179—192.

[10] D.G.Cantor and H.Zassenhaus, A new algorithm for factoring polynomials over finite fields, *Math.Comp.*, 36(1981), 587—592.

[11] H.Cohen and H.W.Lenstra Jr., Primality testing and Jacobi sums, *Math.Comp.*, 42(1984), 297—330.

[12] L.E.Dickson, History of the Theory of Numbers, vol. 1(reprint), Chelsea, New York, 1952 (original publica-

tion 1919).

[13] W.Diffie and M.E.Hellman, New directions in cryptography, *IEEE Trans.Inform.Theory*, 22 (1976),644—654.

[14] J.D.Dixon, Asymptotically fast factorization of integers, *Math.Comp.*, 36(1981), 255—260.

[15] H.E.Dudeney,Canterbury Puzzles(reprint),Dover,New York, 1958 (original first edition published 1907).

[16] P.Erdös, On pseudoprimes and Carmichael numbers, *Publ.Math.Debrecen*, 4(1956), 201—206.

[16a] M.R.Garey and D.S.Johnson, Computers and Intractibility, Freeman, San Francisco, 1979.

[17] C.F.Gauss, Disquisitiones Arithmeticae (transl.A.A. Clarke S.J.), Yale, New Haven, 1966(original first edition 1801).

[18] G.L.Gerver,Factoring large numbers with a quadratic sieve, *Math.Comp.*,41(1983), 287—294.

[19] H.Gunji and D.Arnon,On polynomial factorization over finite fields, *Math.Comp.*, 36(1981), 281—287.

[20] R.K.Guy, How to factor a number, in Proc. 5th Manitoba Conf. on Numerical Math., 1975, pp 49—89.

[20a] K.Ireland and M.Rosen, A Classical Introduction to Modern Number Theory, Springer-Verlag, New York, 1982.

[21] D.E.Knuth, The Art of Computer Programming, Vol. 2(2nd ed.), Addison-Wesley, Reading, Mass., 1981.

[22] M.Kraitchik, Théorie des nombres, vol.2, Gauthier-Villars, Paris, 1926.

[23] A.M.Legendre, Théorie des nombres, vol.1(3rd ed.), Paris, 1830 (1st ed. published 1798).

[24] D.J.Lehman, On primality tests, *SIAM J.Comput.*, 11 (1982), 374—375.

[25] R.S.Lehman, Factoring large integers, ***Math.Comp.***, 28(1974), 637—646.

[26] D.H.Lehmer, An extended theory of Lucas functions, *Ann.of Math.*, (2)31(1930), 419—448.

[27] ——,The sieve problem for all-purpose computers,*Math.Comp.*, 7(1953), 6—14.

[28] ——,Computer technology applied to the theory of numbers, in Studies in Number Theory (W.J.LeVeque, ed.), *MAA Studies in Math.*, 6(1969), pp.117—151.

[29] ——,Strong Carmichael numbers, *J.Austral.Math.Soc. Ser.A*, 21(1976), 508—510.

[30] ——,A history of the sieve process, in A History of Computing in the Twentieth Century (N.Metropolis, J.Howlett and G.-C.Rota, Editors), Academic Press, New York, 1980, pp.445—456.

[31] D.H.Lehmer and R.E.Powers, On factoring large numbers, *Bull.Amer.Math.Soc.*, 37(1931), 770—776.

[32] A.K.Lenstra, H.W.Lenstra, Jr., and L.Lovász, Factoring polynomials with rational coefficients, ***Math. Ann.***, 261(1982), 515—534.

[33] H.W.Lenstra,Jr., Primality testing algorithms[after Adleman,Rumely and Williams], in Séminaire Bourbaki,vol.1980/81,Lecture Notes in Math.901,Springer,Berlin,1981.

[34] G.L.Miller, Riemann's hypothesis and tests for primality,J.Comput.System.Sci.,13(1976), 300—317.

[34a] L.Monier, Evaluation and comparison of two efficient probabilistic testing algorithms, ***Theoret.Comp.Sci.***, 12 (1980), 97—108.

[35] H.L.Montgomery.Topics in Multiplicative Number Theory, Lecture Notes in Math.227,Springer,Berlin,1971.

[36] M.A.Morrison and J.Brillhart, A method of factoring and the factorization of F_7, *Math,Comp.*, 29(1975), 183—205.

[37] D.A.Plaisted, Fast verification, testing and generation of large primes, *Theoret.Comput.Sci*.9(1979)1—16; *errata, ibid.*, 14(1981)345.

[38] S.Pohlig and M.Hellman, An improved algorithm for computing logarithms over GF(p)and its cryptographic significance, *IEEE Trans.Inform.Theory*,24(1978), 106—110.

[39] J.M.Pollard, Theorems on factorization and primality testing, *Proc.Camb.Philos.Soc.*, 76(1974), 521—528.

[40] ——,A Monte Carlo method for factorization, BIT, 15 (1975), 331—334.

[41] C.Pomerance, Recent developments in primality testing, *Math.Intelligencer*, 3(1981), 97—105.

[42] ——,On the distribution of pseudoprimes,*Math.Comp.*, 37(1981), 587—593.

[43] ——,Analysis and comparison of some integer factoring algorithms,in Computational Methods in Number Theory (H.W.Lenstra, Jr., and R.Tijdeman, Editors), Math. Centrum Tract 154(part I),Amsterdam,1982,pp.89—139.

[44] C.Pomerance, J.L.Selfridge and S.S.Wagstaff,Jr.,The pseudoprimes to $25\cdot10^9$,*Math.Comp.*,35(1980),1003—1026.

[45] C.Pomerance and S.S.Wagstaff,Jr.,Implementation of the continued fraction integer factoring algorithm, in Proc.12th Winnipeg Conf.on Numerical Methods and Computing(1982), to appear.

[46] V.R. Pratt,Every prime has a succinct certificate, *SIAM J.Comput.*, 4(1975), 214—220.

[47] M.O.Rabin,Probabilistic algorithms,in Algorithms and

Complexity, New Directions and Recent Results(J.Traub.Editor),Academic Press,New York,1976,pp.21—39.

[48] ——,Probabilistic algorithms in finite fields, *SIAM J. Comput.*, 9(1980), 273—280.

[49] ——,Probabilistic algorithm for primality testing, *J. Number Theory*, 12(1980), 128—138.

[50] R.Rivest, A.Shamir and L.M.Adleman, A method for obtaining digital signatures and public key cryptosystems, *Comm.ACM*, 21(1978), 120—128.

[51] J.B. Rosser and L.Schoonfeld, Approximate formulas for some functions of prime numbers,Illinois *J.Math.*, 6(1962), 64—94.

[52] C.P.Schnorr, Refined analysis and improvements on some factoring algorithms, *J.Algorithms*, 3(1982),101—127.

[53] ——,A Monte Carlo factoring algorithm with finite storage[based on joint work with H.W.Lenstra,Jr.],Seminar on Number Theory, 1981—1982, Exposé #40,Université Bordeaux I ,Telence.

[54] R.J.Schoof,Quadratic fields and factorization,in Computational Methods in Number Theory (H.W.Lenstra, Jr., and R.Tijdeman, Editors), Math. Centrum Tract 155(part I), Amsterdam, 1982,pp.235—286.

[55] A.Shamir, Factoring numbers in $O(\ln n)$arithmetic steps, *Inform.Proc.Letters*, 8(1979), 28—31.

[56] D.Shanks, Class number,a theory of factorization,and genera, in *Proc.Symp.Pure Math.*, vol.20, *Amer.Math. Soc.*,1971, pp.415—440.

[57] ——, Five number theoretic algorithms, in Proc.2nd Manitoba Conf.on Numerical Math.(1972).

[58] ——,The infrastructure of a real quadratic field and

applications, in Proc. 1972 Number Theory Conf., Boulder, Colorado, pp. 217—224.

[59] ——, Square-form factorization, a simple $O(N^{1/4})$ algorithm, to appear.

[60] R. Solovay and V. Strassen, A fast Monte-Carlo test for primality, *SIAM J. Comput.*, 6(1977), 84—85; erratum, ibid., 7(1978) 118.

[61] I. M. Vinogradov, Elements of Number Theory (rev. 5th ed.), Dover, New York, 1954.

[62] H. C. Williams, Primality testing on a computer, *Ars Combin.*, 5(1978), 127—185.

[63] ——, A p+1 method of factoring, *Math. Comp.*, 39(1982), 225—234.

[64] M. C. Wunderlich, A running time analysis of Brillhart's continued fraction factoring method, in Number Theory, Carbondale 1979. Lecture Notes in Math. 751, Springer, Berlin, 1979, pp. 328—342.

（张明尧编译，潘承彪校）

有 限 集

中学生对有限集的认识是很直观的：利用正整数1,2,3,…，可把这个集数到最后一个元．初看起来，有限集概念相当简单，但如果深入研究下去，却非常有趣，而且还是大学数学专业学习的内容，是进一步学习集合论思想和方法的入门．

我们准备用标准的集合论观点来讨论有限集①．为此，有必要粗略地描述下面用到的几个集合论概念：

两个集合（不一定是有限集）相等当且仅当它们有相同元素，即集合 $A=B$，当且仅当其中任一个必是另一个的子集．如果存在A到B上的一一映射F（称为双射），则称A与B等价或基数相等，记为F：$A\approx B$．

集合 $\bar{n}=\{0,1,2,\cdots,n\}$称为自然数集② $\bar{n}$，所有自然数的集合记为N．

导出有限集性质的主要工具是数学归纳法：如果A是N的子集，$0\in A$，且对每个n，只要$n\in A$，便有$(n+1)\in A$，则 $A=N$．数学归纳法的另一形式是最小数原理：自然数集的每个非空子集有最小元．

定义1 如果集合A等价于某个自然数集 $\bar{n}$，则A是有

① 假定读者对集合的概念，子集，集合的运算，并$(A\cup B)$，交$(A\cap B)$，差$(A-B)$都是知道的．

② 这里把0称作自然数．

限集，否则A是无限集。

利用数学归纳法，易证有限集有下列性质：

(1) 如果A,B是有限集，则并集$A\cup B$也是有限集。

(2) 有限集的每个子集是有限集。

(3) 如果A是有限集，则对每个集B，交集$A\cap B$，差集$A-B$都是有限集。

(4) 如果A是有限集，则A的所有子集组成的集$\mathscr{P}(A)$是有限集。

(5) 如果A是有限集，$B\subseteq A$，且$A\approx B$，则$A=B$。

(6) N的每个非空有限集有最大元。

为后面的需要，现给出偏序，线序，良序，集合的极大元、极小元，最大元、最小元的定义：

(1) 在集合A的元素之间引进二元关系① R，它具有这样的性质：当$a,b,c\in A$时，有

(i) aRa,

(ii) 若aRb,bRc,则aRc,

(iii) 若aRb,bRa，则$a=b$,

那么称A关于二元关系R成偏序集。例如，任意集合关于包含关系$\subseteq$是偏序集。

(2) 集上（关于某二元关系R）的极大元与最大元，极小元与最小元是不同的概念（严格定义见定义2,定义3）。极大元是这样的元素：(在二元关系R的意义下)没有再“大”

---

① 即在集合A中对其元素给出一种关系，对任意两个元素a,b，可判断a和b（与次序有关）有没有这种关系R。若有这种关系，就记作aRb。如实数集合中的大于（$>$），等于（$=$），小于（$<$），不小于（$\geqslant$），不大于（$\leqslant$）等和整数集合中的整除关系。

于它的元，而最大元"大"于任何其它元；极小元与最小元的区别相类似。例如，设 $A=\{\{1\},\{2\},\{1,3\}\}$，则 $\{1,3\}$ 关于子集包含关系是极大元，但不是最大元，因 $\{2\}$ 不是它的子集。显见，最大元总是极大元。

定义2 设在集合 A 中给出了一个二元关系 R，且 $a\in A$，如果对任意 $b\in A$，从 aRb 可推出 $b=a$，称 a 是 A 的 R-极大元。

定义3 设在集合 A 中给出了一个二元关系 R，且 $a\in A$，如果对任意 $b\in A$，从 bRa，可推出 $a=b$，称 a 是 A 的 R-极小元。

定义4 设在集合 A 中给出了一个二元关系 R，且 $a\in A$，如果对任意 $b\in A$，有 bRa 或 $b=a$，称 a 是 A 的 R-最大元。

定义5 设在集合 A 中给出了一个二元关系 R，且 $a\in A$，如果对任意 $b\in A$，有 aRb 或 $a=b$，称 a 是 A 的 R-最小元。

(3) 集 A 关于二元关系 R 是线序的，如果

(i) A 关于 R 是偏序的，

(ii) A 关于 R 是连通的，即对任意 $a,b\in A$，aRb 与 bRa 至少有一个成立。

例如，有理数集关于二元关系 $\leqslant$ 是线序的。

(4) 集 A 关于二元关系 R 是良序集，如果

(i) A 关于 R 是线序的，

(ii) A 的每个非空子集有 R-最小元。

例如，N 关于二元关系 $\leqslant$ 成良序集。

因为有限集概念比较直观简单，自然希望在定义和导出它的性质时，不要用过强的工具，如无穷公理。二十世纪初，波兰数学家阿尔费雷德、塔尔斯基不依赖无穷公理建立了有限集理论。下面给出几种描述有限集的方法。

定理1 A 是有限集当且仅当 $\mathscr{P}(A)$ 的每个非空子集对

包含关系$\subseteq$有极大（或）极小元，即有$\subseteq$-极大（或$\subseteq$-极小）元。

证 由有限集性质（4），A的所有子集组成的集$\mathscr{P}(A)$是有限集。设S是$\mathscr{P}(A)$的一个非空子集，由性质（2)知，S是有限集；因$S\subseteq\mathscr{P}(A)$，S的每个元是A的子集；由性质（2）知，A的每个子集是有限集，故S的每个元是有限集。对每个$a\in S$，用$|a|$表示集合a中元素的个数，设$y=\{|a|,a\in S\}$，则y是自然数的有限集，由性质(6)，y有最大元，设为m；并设$b\in S$，$|b|=m$，则b一定是S的$\subseteq$-极大元.否则,若b是某个元$c\in S$的真子集,则$m=|a|<|c|$，与m的定义矛盾。因此，不可能有$c\in S$使$m<|c|$。类似可证S一定有$\subseteq$-极小元。

反之，设$\mathscr{P}(A)$的每个非空子集有极大元，但A不是有限集，则可以构造$\mathscr{P}(A)$的一个非空子集，它没有$\subseteq$-极大元，从而得出矛盾。设$E(A)$是A的所有有限子集组成的集：$E(A)=\{a: a\subset A,|a|\in N\}$。我们断言，$E(A)$无$\subseteq$-极大元。若设$b$是$E(A)$的$\subseteq$-极大元，因$b$是有限集，而$A$不是，则$A-b\neq\varnothing$。设$a\in A-b$，集$\{a\}$是有限集，由性质（1），$b\cup\{a\}$是有限集，且是$A$的子集，故$b\cup\{a\}\in E(A)$，$b$是$b\cup\{a\}$的真子集，这与$b$是$E(A)$的$\subseteq$-极大元矛盾。因此，$E(A)$无$\subseteq$-极大元，这与定理条件矛盾。故$A$是有限集。对必有极小元的情形可类似证明。

我们将用“R^{-1}”表示关系R的逆，即aRb当且仅当$bR^{-1}a$。

定理2 **A是非空有限集当且仅当存在关系R，使A关于R^{-1}与R都是良序集。**

证 设A是有限集，由定义1，存在自然数集$\bar{n}$，使$A\approx\bar{n}$，即A有$n+1$个元，于是A可表示成$A=\{a_0,a_1,a_2,\cdots,a_n\}$。我们定义$A$上的关系$R$，使对每个$i,j=1,2,\cdots,n$，当且仅当$i\leqslant j$时，$a_iRa_j$。由于$R^{-1}$-最小元即为$R$-最大元，故$R$与$R^{-1}$均使$A$成良序集。

反之，设有A上的关系R，使A关于R与R^{-1}均成良序集，我们用反证法证明A是有限集。不然的话，我们将证明，A关于R是良序集时，存在A的子集没有R^{-1}-最小元。假若A是无限集，对A的每个有限子集a，$A-a\neq\varnothing$，由良序集的定义，$A-a$有R-最小元。利用数学归纳法，可以定义从N到A的映射F：

$F(0)=A$的R-最小元，

$F(1)=A-\{F(0)\}$的R-最小元，

$F(2)=A-\{F(0),F(1)\}$的R-最小元，

$F(3)=A-\{F(0),F(1),F(2)\}$的$R$-最小元，

…………………………………………………………。

设映射的像集是B，则$B\subseteq A$，可以断言，B没有R^{-1}-最小元，否则，设b是B的R^{-1}-最小元，则$b\in B$，故有$n\in N$，使$b=F(n)$。由F的定义，$F(n)RF(n+1)$，且$F(n)\neq F(n+1)$，即$F(n+1)R^{-1}b$和$F(n+1)\neq b$，这与b是B的R^{-1}-最小元矛盾，故B不可能有R^{-1}-最小元，即A关于R^{-1}不是良序集，与条件矛盾。可见A是有限集。

定理3 A是非空有限集当且仅当A可以线序，且A的每个线序是良序。

证 有限集A可以线序的证明与定理2证明的第一部分相同。现设关系R赋A线序（即A关于关系R是线序的），

下面证明R赋A良序。还是用反证法，如果R不赋A良序，一定存在A的非空有限子集，它没有R-最小元，我们用最小元原理证明这一点不可能。设B是A的非空子集，它具有性质：

(i) 没有R-最小元，

(ii) A的每个比B元素少的非空子集均有R-最小元。

设b是B的一个元，由假设，B非空，故$B-\{b\}$非空（为什么?），且元素少于B，由性质 (ii)，$B-\{b\}$有最小元，比如c。于是有

(iii) 对一切$s\in B-\{b\}$，cRs。因B关于R连通，故又有

(iv) bRc或cRb。

如果cRb，则由 (iii)，c是B的R-最小元，这与 (i) 矛盾；如果bRc，因关系R可传递，由性质(iii)，bRs对一切$B-\{b\}$成立，则b是A的R-最小元，也与 (i) 矛盾。于是，A不可能有无R-最小元的子集，即A关于R成良序。

现在证明充分性。假定A可以线序，且A的每个线序都是良序，并设赋A线序的关系为R。如能证明R^{-1}赋A线序，由假设，R与R^{-1}都赋A良序，根据定理2，可知A是有限集。注意到$aR^{-1}b$当且仅当bRa，由线序定义易证，如果R赋A线序，则R^{-1}也赋A线序。这就证明了定理。

对下述定理只给予略证，请读者补充证明细节。

定理4 A是有限集当且仅当$\mathscr{P}(A)$是唯一满足下列条件的集S：

(i) $S\subseteq\mathscr{P}(A)$，

(ii) $\varnothing\in S$，

(iii) 若$a\in A$，则$\{a\}\in S$，

(iv) 如果$b\in S, c\in S$，则$b\cup c\in S$。

证 设A是有限集，B是A的任一子集，则B是有限集，于是存在自然数n，有

$$B=\{a_1, a_2, \cdots, a_n\}.$$

利用性质 (ii)—(iv)和关于n的数学归纳法，可证$B\in S$，由B的任意性，可得 (v)：$\mathscr{P}(A)\subseteq S$，结合 (i)，便证得$S=\mathscr{P}(A)$。

反之，设$\mathscr{P}(A)$是唯一满足四个条件的集合，然后证明，A的一切有限子集成的集$E(A)$满足四个条件，于是必有$E(A)=\mathscr{P}(A)$，即得，A的每个子集是有限集，故A是有限集。

定理5 A是有限集当且仅当存在A到A的映射F，使$F(A)=A$，但不存在A的真子集A^1，使$F(A^1)\subseteq A^1$。

证 设A是有限集，$A\neq\varnothing$，则有自然数n，使$A=\{a_1, a_2, \cdots, a_n\}$。定义$A$上的映射$F$：$F(a_i)=a_{i+1}(i=1,2,\cdots,n-1)$，而$F(a_n)=a_1$。如果$A=\varnothing$，令$F=\varnothing$，则$F$就是定理所要求的映射。

反之，设F是满足定理条件的映射，若$A=\varnothing$，显然A是有限集。现设$A\neq\varnothing$，且$a\in A$，定义以N为定义域的映射G：

$$G(0)=a,\quad G(n+1)=F(G(n)),\ n\in N,$$

即

$$G(0)=a,$$

$$G(1)=F(a),$$

$$G(2)=F(F(a)),$$

$$G(3)=F(F(F(a))),$$

••••••••••••••••••••••••

设B是G的像集，因F映A到自身，故$B\subseteq A$。显见，$F(B)\subseteq B$，因而，由条件知，B不是A的真子集，所以，只有$B=A$。如果没有自然数$n>0$，使$G(n)=a$，则$A-\{a\}$是A的真子集，由$A=B$知，$F(A-\{a\})\subseteq A-\{a\}$。由条件知，这是不可能的。设$m$是使$G(n)=G(0)$的最小自然数，即$m$是集$\{n:n>0$，且$G(n)=G(0)\}$中最小的自然数，于是$A\subseteq\{G(0), G(1), \cdots, G(m-1)\}$。可是，$A$是有限集的子集，当然是有限集。

根据有限集性质（5），有限集不能与它的真子集等价。二十世纪初，R.Dedekind提出，可用这性质定义有限集，我们称具有这性质的集为Dedekind有限集。

定义6 A是Dedekind有限集，如果它不等价于它的任何真子集。

由性质（5），每个有限集是Dedekind有限集。保罗·柯恩在最近的研究表明，添上选择公理①，才能证明其逆。由于Dedekind有限集概念本身就很有趣。这里给出它的一些等价形式。集A称为可列无限集，当且仅当$N\approx A$。

定理6 A是Dedekind有限集当且仅当A不含可列无限子集。

证明 设A包含可列无限子集B，则$N\approx B$。假定F是以N为定义域，B为像集的双射，

$$F:N\approx B$$

且G是以A为定义域的映射，定义为：

---

① 选择公理可参看聂灵沼和丁石孙合著的《代数学引论》，高等教育出版社，1989。

$$G(u)=\begin{cases}u, & u\in A-B,\\ F(n+1), & u\in B\text{且}u=F(n),\end{cases}$$

于是，G是$A-B$上的恒等映射，G把B上的每个元素变换成下一个元，因此G是一一映射．G的像集是$A-\{F(0)\}$，它是A的真子集．因此，A等价于它的真子集．故A不是Dedekind有限集．可见，如果A是Dedekind有限集，就不含可列无限子集．

反之，设A不是Dedekind有限集．则有A的真子集B和双射F，使$F:A\approx B$．设$a\in A-B$，如定理5的证明一样，利用数学归纳法，可构造以N为定义域的映射G：

$$G(0)=a=F^{(0)}(a),$$

$$G(1)=F(a)=F^{(1)}(a),$$

$$G(2)=F(F(a))=F^{(2)}(a),$$

$$G(3)=F(F(F(a)))=F^{(3)}(a),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

可以肯定，G是一一映射．因若有自然数$m,n,m\leqslant n$，使$G(m)=G(n)$，则$F^{(m)}(a)=F^{(n)}(a)$．但F是一一映射，应用F的逆F^{-1}作用于上式两端m次，便有$a=F^{(n-m)}(a)$．但$a\in A-B$，如果$n-m>0$，则$F^{(n-m)}(a)\in B$，上式只有当$n=m$时才成立．因此，G是映N到A的一一映射，G的像集是A的可列无限子集．

下面定理中用到两个符号："$\prec$"与"X^+"．X表示集合，X^+表示比X多一个元素的集合，而$A\prec B$表示存在A到B的一一映射，但$A\approx B$不成立．

定理7 A是Dedekind有限集，当且仅当$A\prec A^+$．

证 首先，如果A是Dedekind有限集，则A^+也是，显然

$A \subset A^{+}$，故恒等映射是映 A 到 A^{+} 的一一映射。如果 $A \approx A^{+}$，则 A^{+} 等价于它的真子集，与 A^{+} 是 Dedekind 有限集矛盾，因此 $A \prec A^{+}$。

反之，如果 A 不是 Dedekind 有限集，由定理 6，A 有可列无限子集 B，设 F 是映射，使

$$F: N \approx B.$$

可定义映射 G，以 $A^{+} = A \cup \{c\}$ 为定义域，且使

$$G(c) = F(0),$$

$$G(F(n)) = F(n+1), \quad n \in N,$$

$$G(u) = u, \quad u \in A - B$$

易见，$G: A^{+} \approx A$，即 $A \prec A^{+}$ 不成立。

定理8 A 是有限集当且仅当 $\mathscr{P}(\mathscr{P}(A))$ 是 Dedekind 有限集。

证 如果 A 是有限集，由性质（4），$\mathscr{P}(\mathscr{P}(A))$ 是有限集，由性质（5），每个有限集是Dedekind有限集。

反之，设 A 不是有限集。对每个 $n \in N$，用 S_n 表示 A 的所有含 n 个元的子集成的集，即

$$S_0 = \{\varnothing\},$$

$$S_1 = \{\{a\}, a \in A\},$$

$$S_2 = \{\{a, b\}, a, b \in A, \text{且} a \neq b\},$$

$$\cdots\cdots$$

因 A 不是有限集，故对一切 $n \in N$，$S_n \neq \varnothing$，因此，对一切 $n, m \in N, n \neq m$，有 $S_n \neq S_m$；再有，对每个 $n \in N$，$S_n \in \mathscr{P}(\mathscr{P}(A))$，故集 $\{S_n, n \in N\}$ 是 $\mathscr{P}(\mathscr{P}(A))$ 的可列无限子集。由定理6，$\mathscr{P}(\mathscr{P}(A))$ 不是Dedekind有限集。证毕。

（徐平五编译，潘承彪校）

不能证明的命题

（不变量的运用）

人们常说，只要你刻苦努力和机敏，一切困难都能克服。在这里，我们必须正确理解这句话的含意。世界上有许多事情是能办到的，但有些事情是办不到的。那末，如何阐明这一点呢？实际上，对于一些被认为是不可能完成的事情中，会发生两种不同的格局（注意，这里指的是理论范畴）。

（Ⅰ）在某些时候，有人认为某件事是不能办到的，但实际上是能办到的。

（Ⅱ）在某些时候，某件事确实可以证明不能办到。当然，即使如此，也还可能出现下述希望：对原先的约定作些修改，情况就起了变化，并由此而达到期望的结果。

显然，前者只说明一个错误的判断；而后者正是表达了我们所说的“数学中的不可能性”。

我国古代有一个传说，讲的是一个出售矛和盾的人，自称这种盾可以抵挡任何锋利的矛，而他的矛则又是无坚不穿的。很明显，这是不能成立的，只要以其矛试其盾即可。这就是数学家们在证明某些命题是不可能成立时使用的方法。他们指出，相反的假设将导致谬误和矛盾。当然，必须强调的是，数学家们在这里讨论的是严格定式化了的问题。因此，事情一旦被证明为不可能，它就不是第一种格局，除非

改变前提。

这样，“不能成立的命题”似乎没有什么意义了。然而，无论怎么讲，它们属于数学中最令人惊叹的结果，因为它们具有某种神秘的力量。大家知道，数学已经（在一般水平上）证明人们不能仅用直尺与圆规三等分一个任意角(当然，这并不妨碍工程师们在需要时去三等分一个角)。现在，让我们反过来探讨一下。如果三等分一个任意角是可能的，那么，如同二等分一个角，极易想象它大概也是用若干步骤把它构造出来，只不过更复杂些而已。因此，这种不可能性的证明之神秘性就在于：在众多错综复杂（可能包含上千个作图步骤）的构造证明中我们如何知道它是不正确的呢！看来，在这里显然须有某种普遍适用的准则来作出判断。本文的主要部分就是要举例说明在各种全然不同的背景下如何运用某种原理来探求“不可能性”的证明。我们列举六个例子，贯穿其中的共同线索就是所谓不变量原理。不变量的严格定义属于逻辑和哲学范畴。通俗地讲，当一位研究者面临这样一种局面，要完全分析被研究对象的现状是十分复杂的，于是他就去寻求某种特定的事物，所谓不变量。这样一来，使现状的共性具体化了，而且这种不变量有希望被容易地计算出来。研究者接着指出，在系统中的某些变换下，这种不变量是不能改变的（这正是名称的由来）。因此，如果“求解”这个确定的问题，是要改变不变量，那末这种解是不可能获得的。

不变量的例子在许多学科中均可找到，例如能量，动量，角动量是标准的物理不变量。扩大一点说，人的指纹就是解剖学中的不变量。在数学中，奇偶性以及长度、面积等数量

也是常见的不变量。值得再次强调的是，正是某些不变量确立了“不可能证明的命题”，然而它们是如此明显致使有时逃离了我们的注意力。在下面论述的例子中，不变量并非明显地摆在我们面前，而是需要去揭示。

例1（棋走顶角） 设想在一个画有方格的棋盘(见图1)上，有一个棋子可在方格之间走动，每步走一方格，或水平方向或垂直方向（不许走斜线，也不许越两格）。开始时棋子处于棋盘左上角位置（见图1），现在要求它走过棋盘上所有方格，但每个方格又只许走过一次，最后落到右下角的“×”位置上。证明这是不可能的。（例1—例4的解答将在例4后给出）

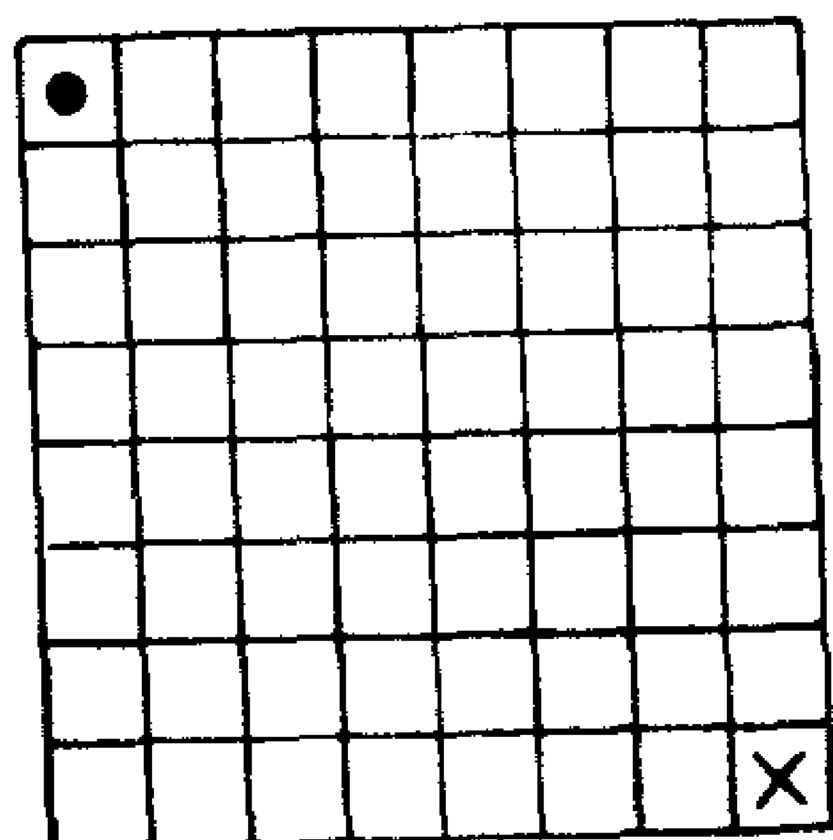


图 1

例2（15-迷宫） 设有一个画有4×4的方格盘，以及15个与方格大小相同的方块，其上编有从1到15的号码。将它们放于盘中，其中有一格是空着的（见图2）。现在规定这些方块都可以沿水平或垂直方向移动到相邻的空位上(不许把方块移出盘外)。证明依照这一移动规则，不可能使图2(b)的位形变成图2(a)的位形。

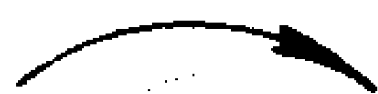
1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	

图2(a)　希望到达的最后位形

1	2	3	4
5	6	7	8
9	10	11	12
13	15	14	

图2(b)　这里的14与15是错位的

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	



1	2	3	4
5	6	7	8
9	10	11	
13	14	15	12

图2(c)　举例，方块12移向空位

例8（马步）　在一块3×3见方的棋盘的4个角位上放有四只(棋)马，左下角是白色的，右下角是黑色的，左上角和右上角是红色的(图3)．这四只马的跳法与象棋中马的跳法一

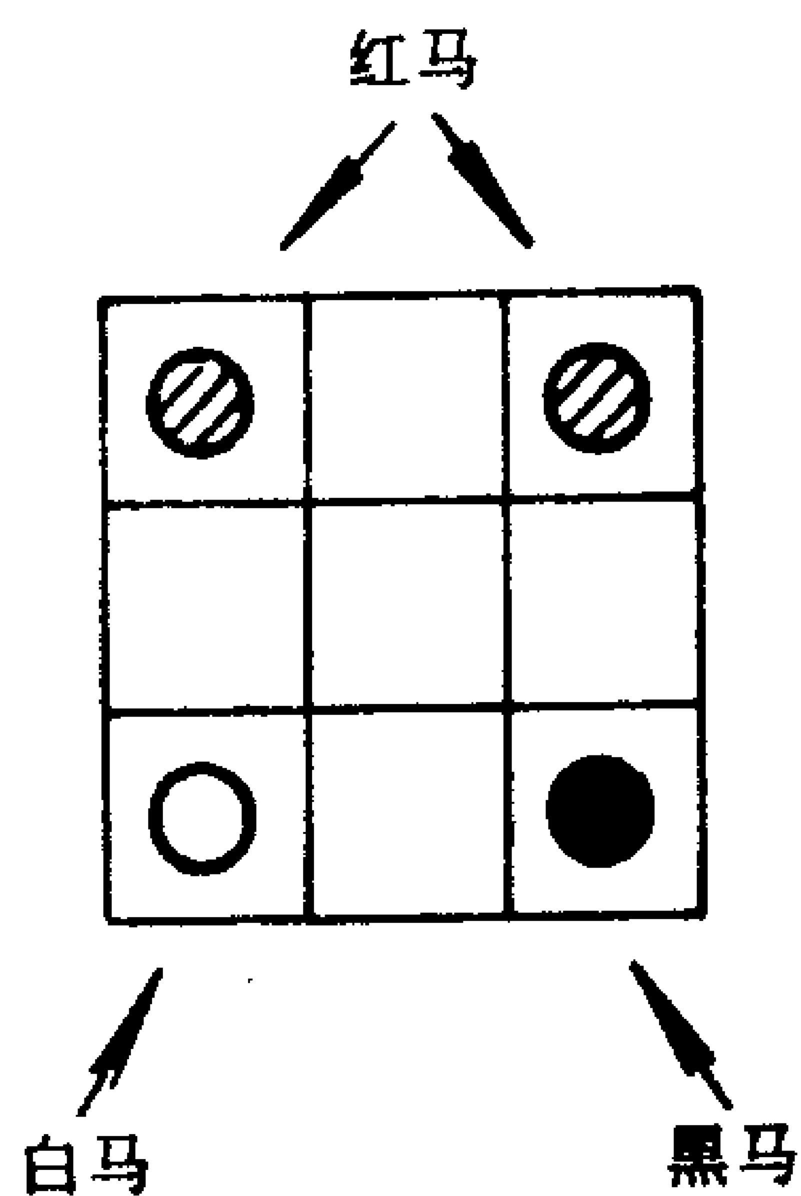
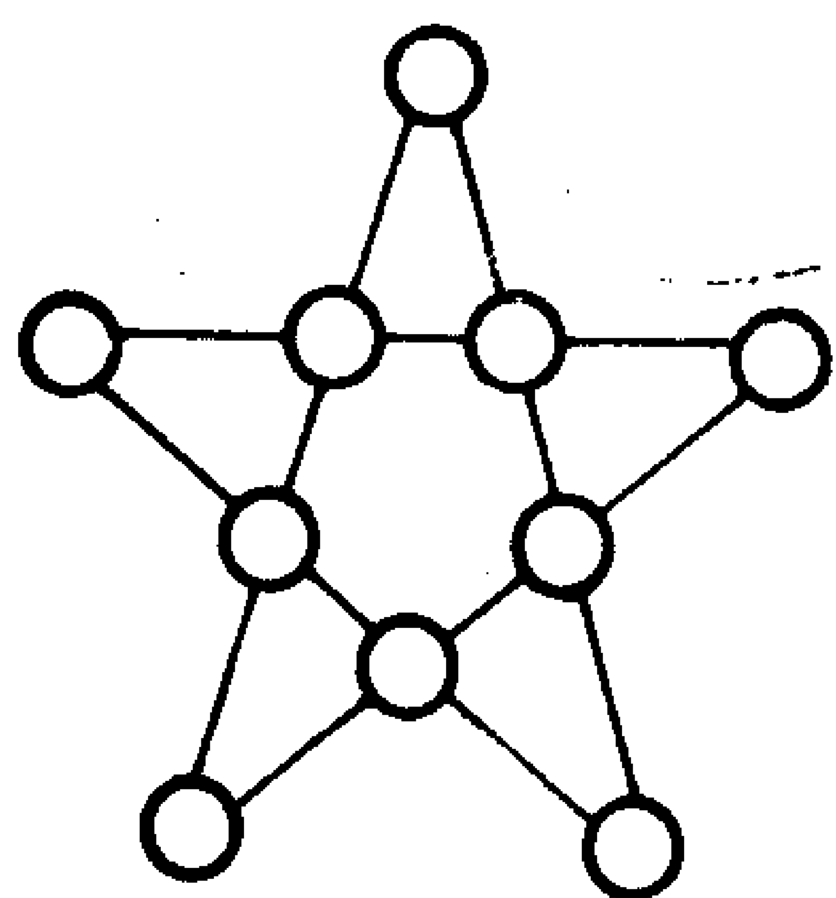


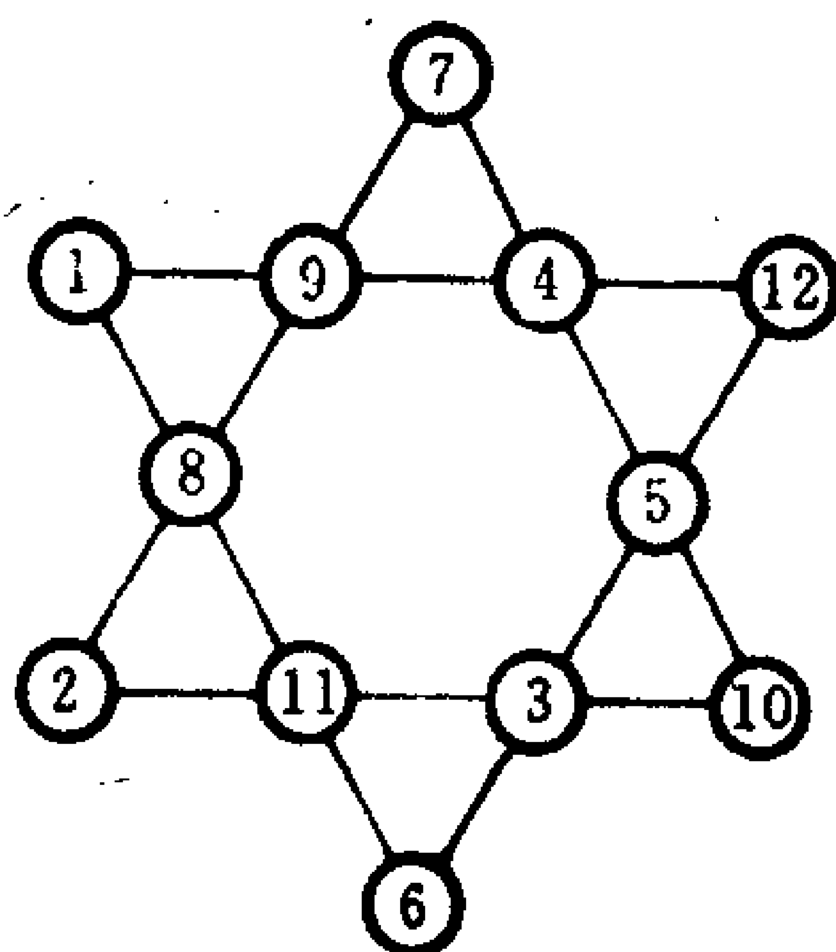
图　3

样，当然不能越出此方块，并且不能吃掉对方(两只马不能同时位于同一方格中)．证明白马与黑马不可能互相交换位置．也就是说，不能使白马跳到右下角，而同时黑马位于左下角(不要求红马处于原位)．

例 4 (五角魔星)　一个五角星(图4(a))有10个交叉点，证明不可能把从1到10这10个数字写在这10个交点处(每一个位置写一个数)，使得这五角星中任一直线上的四个数之和均相同．(六角魔星如图4(b)所示)



(a)



(b)

图 4

例1—例4的解答．

例1的解答　如果把棋盘画成黑白相间的方格(图5)，那末该棋每走一步是从白方格走到黑方格，或者相反．从而要使它从开始位置(白方格)走到终点位置(另一白方格)必须走动偶数次．然而，由于第一方格已被占位，故只有63个方格尚需要占位，它是一个奇数．这一矛盾证明了此命题的不可能性．

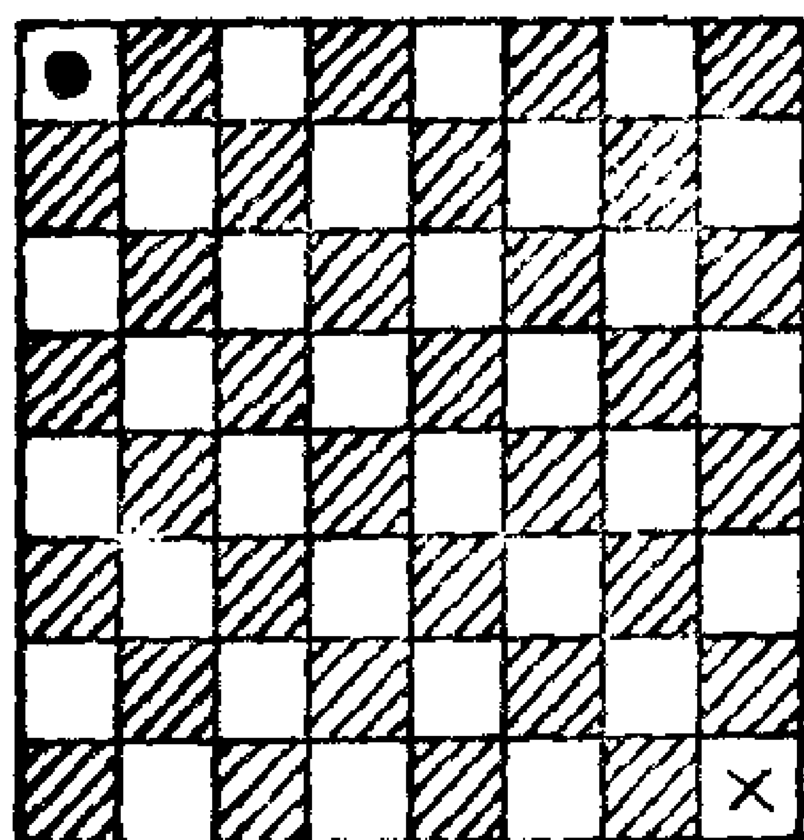


图 5

例2的解答 在这里，我们采用与例1相同的方法，不过再加上一个新的看法。首先，方块每完成一次移动相当于该空格依水平或垂直方向移动一格，就像例1中的走棋（见图6）。于是，根据例1中所述的同样理由，我们必须移动偶

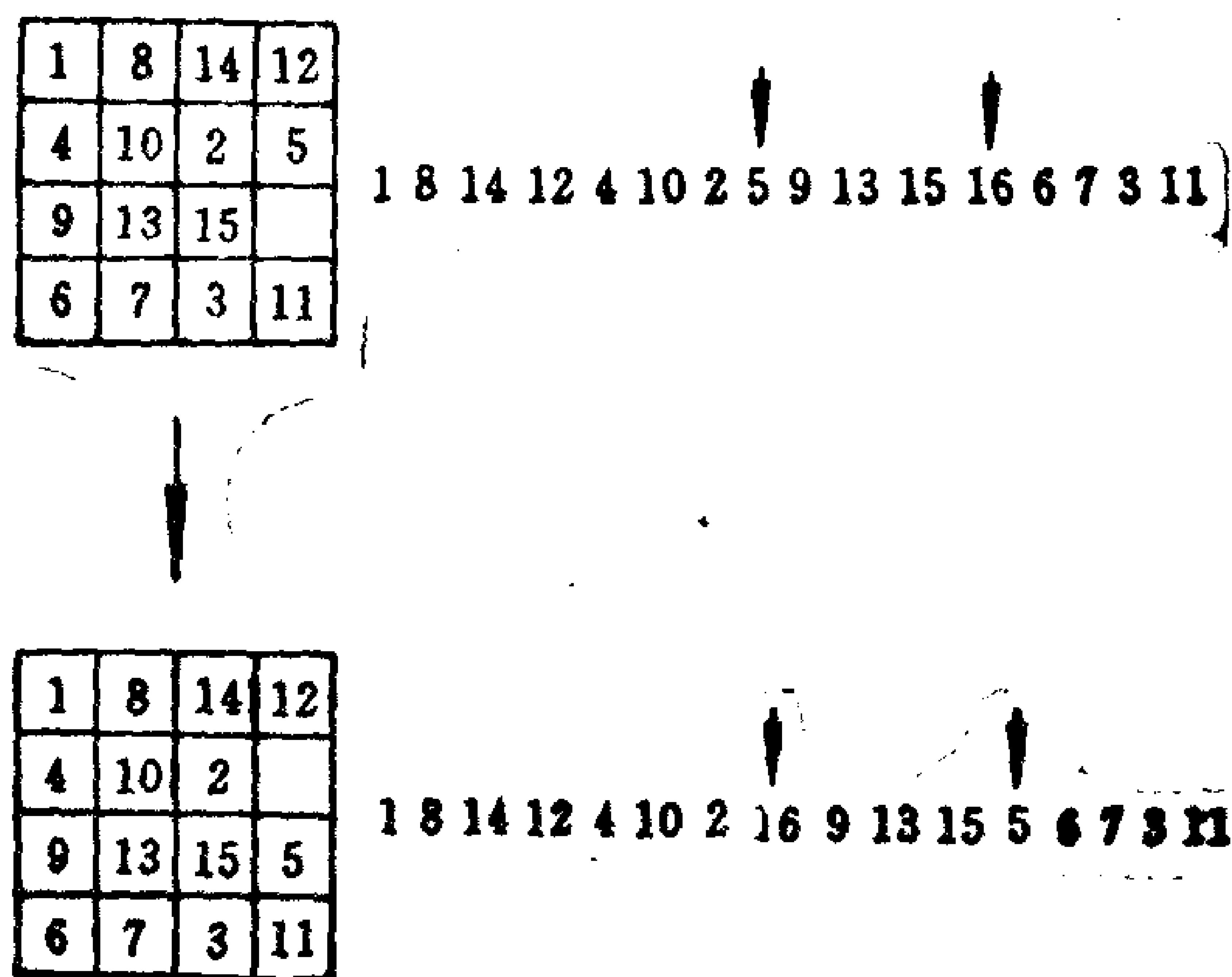


图6 移动示例

数次才能使空格回到它的原位上。然而，可以证明调换第14——方块与第15——方块(如果能做到)，需要奇数次移动，从而我们又遇到如同例1中所述的同一矛盾。

为了证明移动的次数是奇数，需要用到群论中的置换理论，在这里只能说明一个大意，如果我们把空方格记上数16，那末这15个方块加上空方格一起的任意一种位置状态就相当于一个排列。这一排列是由图2(a)所表的自然排列经过一个置换而得到的。从而现在的问题是要从排列(A)——图2(b)位形置换为(B)——图2(c)位形：

(A)

1,2,3,4,5,6,7,8,9,10,11,12,13,15,14,16。

(B)

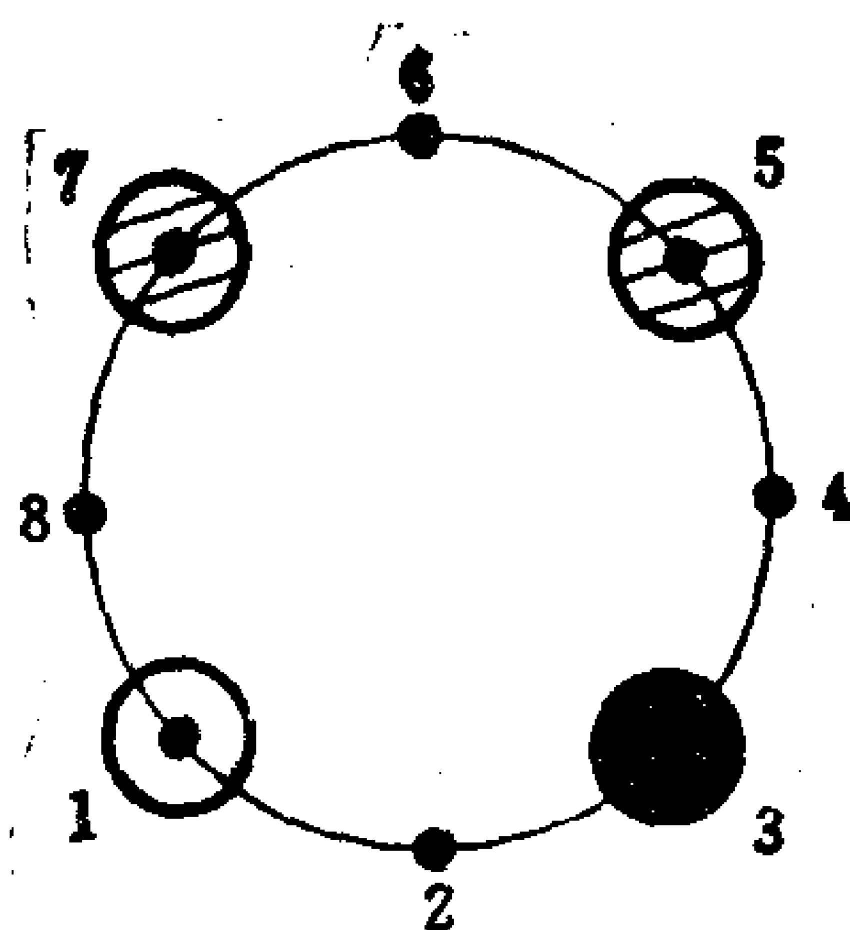
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16。

显然，从(A)到(B)可以通过1次对换，即14与15对换而得到；当然也可以通过16与15对换，再通过16与14对换，最后作16与15对换等3次对换而得到。不过，无论是以何种方式进行各种对换，可以证明其对换次数总是奇数(奇置换)这一点是不为改变的。

例3的解答 在3×3棋盘的中心方格是任一只马都跳不到的位置，其它八个位置如图7(a)所示。因此，马棋可能的跳位只是1→2→3→4→5→6→7→8→1，以及返回的路(即上述路径的逆：1→8→7→6→5→4→3→2→1)。从而我们可以将这一命题用一个圆周上的8个点来表示(图7(b))。在该圆周示意图中，可允许的移动就只是在圆周上简单地按顺时针或逆时针方向从一个位置到相邻的一个位置移动，当然不能两点相重于一个位置，也不能越过另一个点。这样，解此问

7	2	5
4		8
1	6	3

(a)



(b)

图 7

题的不变量看起来应是拓扑性质的。因为，为了交换白位与黑位的位置，两个红位最后必须同时占据数2的位置，这违反了规定。

例4的解答 我们的证明基于下述思想：从1到10这一数集太小了，以致不能用中间那些“适度”的数来平衡数1与数10这两个“极端的数。因此在这里，大小比奇偶起着更关键的作用(另一种用奇偶的解法可参见[1])。我们有下面结论：

(i) 如果有一个解存在，那末在每条直线上的数的和将是22(等于数1→10的平均数5.5之4倍)。

(ii) 如果有一个解存在，那末数1与数10一定位于一条直线上。

证明 若不然，则取六个其它的数放在通过数1的两直线上：此六个数的和$\leqslant 9+8+7+6+5+4=39$。但依(i)，它们的和应为$21+21=42$。

注 图8是某种位形示意，星中数1与数10的特定位置

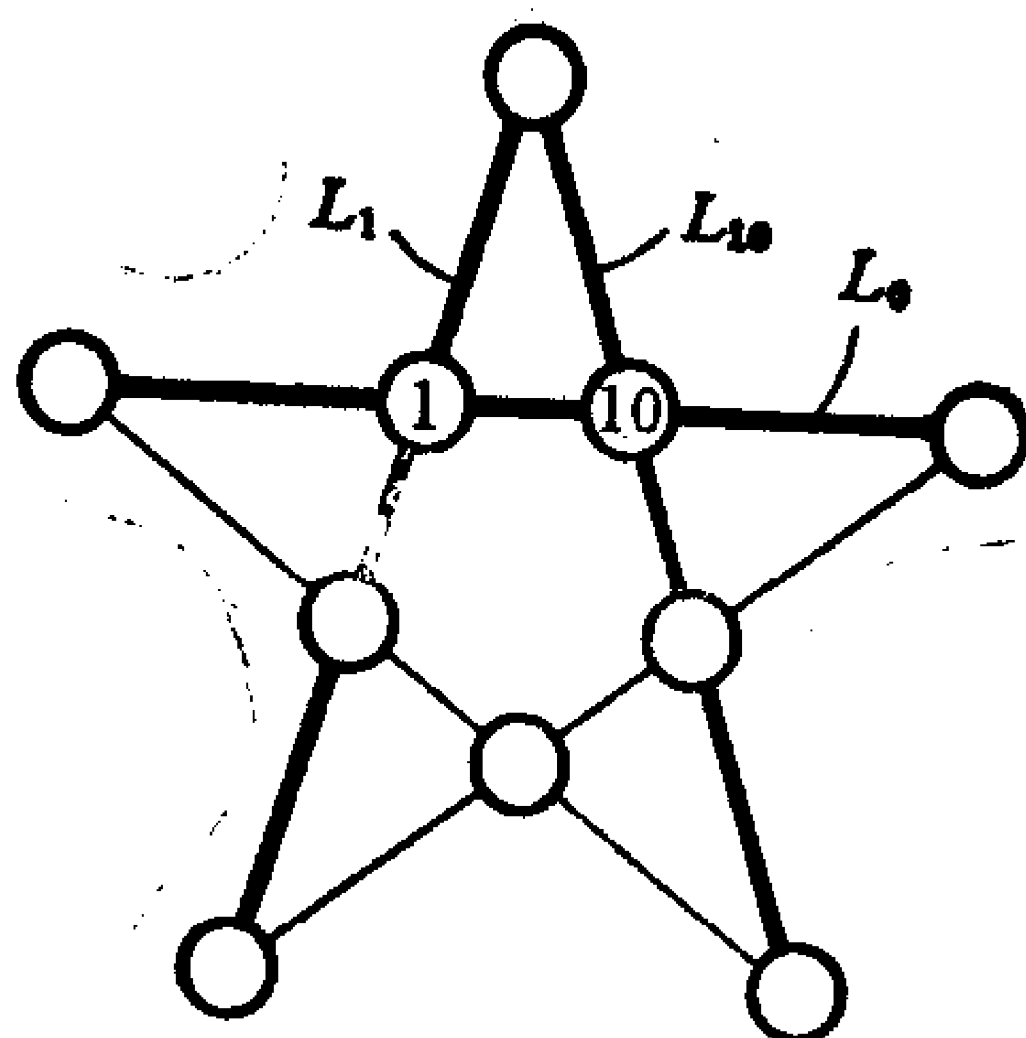


图 8　五角星，这里数 1 和数10处于特定的位置，并标明了三条线：L_0，L_1，L_{10}

与证明无关．从(ii)知，数1与数10位于一条直线上．我们的证明只是要求五角星中任两条直线相交．

(iii) 现在假定数1与数10位于一条直线上，记为 L_0（图 8），而其它两条通过数1和通过数10的直线记为 L_1，L_{10}．

我们必须计算有多少种放置这些数(除1与10外)的方法？还好，数目不大．它们的所有可能之组合是：L_0上两个空位或L_1以及L_{10}上的各三个空位(记住，在每条直线上的四个数之和是22)．

L_0	L_1	L_{10}
(除 1，10外)	(除 1 外)	(除10外)
9，2	8，7，6	3，4，5
8，3	9，7，5	2，4，6
7，4	无	无
6，5	9，8，4	2，3，7

此表表示各种可能的数的组合填入如图8所示的直线L_0，L_1和L_{10}上的空位．四个数的和必须是22．

于是，在表中的第一种情形，直线L_0,L_1和L_{10}将包含下述四个数组：

$L_0=\{1,10,9,2\}$，$L_1=\{1,8,7,6\}$，$L_{10}=\{10,3,4,5\}$。

可这样的五角星中，直线L_1与L_{10}将无公共元素，但这是不可能的。因为五角星中任意两条直线必有公共点。在表中其它的两种情形也是这样。这一矛盾证明了构成五角魔星的不可能性。

议论 在这些证明中究竟用了哪些不变量？在例1和例2里，显然是奇与偶；例3是"中间性"的拓扑概念；例4则是数目大小的限制。当然，在每一个问题中，还有构思（或妙计），使这些不变量发挥功效。它们是：

棋盘上方格的彩色化(例1)。

奇置换与偶置换的概念，以及命题：要求奇数次移动把某地挪到另一地(例2)。

把棋盘上的马的跳步设计到圆周上的移动(例3)。

把注意力集中于两个极端数字1和10的方法(例4)。当然，在这里还用到一个事实：五角星中任意两条直线都相交，它说明了另一种不变量。

在下面要介绍的问题中，不变量就比较复杂，且要依靠伽罗瓦理论才能阐明。在这里，只能简略地来介绍，不过，在陈述这些问题时，我们将作一些评注，也许对不知道伽罗瓦理论的读者会有所启发。

例5（三等分一个任意角） 利用希腊几何的"古典工具"——直尺(无标记的)与圆规，能否把任意的一个角三等分？判定这样一个问题是否有解？二千年来一直没有解决。最后，这一命题被证明为不可能时，已到1800年左右了(顺便

提一下，这个结果似乎呈现出“大众性定理”的状态，至少不能完全归功于任何一个人)。P.Erdös 甚至指出，古代的阿基米德就猜到这一问题是不可解的，而且已悟到(在那个时代)他无法证明这一点。因此他集中精力于自己能解决的许多创造性工作中。

如果真是如此，那末阿基米德的推断是正确的。因为到现在为止，关于三等分任意角问题的不可解性之所有证明都涉及到所谓《抽象代数》中的一些思想。这是在十九世纪以前尚未发展起来的一门学问。实际上，真正用“几何”所做的事是极小的，而转折点则是把几何问题化为代数问题。

让我们简单地分析一下，如果给定一个角度 θ(单位长度已给定)：$0<\theta<\frac{\pi}{2}$，我们就可以求出 $a=\cos\theta$，假定能够求出 $x=\cos\frac{\theta}{3}$，也就可以作出 $\frac{\theta}{3}$ 的角度。应用公式

$$\cos\theta = 4\cos^3\frac{\theta}{3} - 3\cos\frac{\theta}{3},$$

可知 x 满足方程

$$4t^3 - 3t - a = 0.$$

设

$$f(t) = 4t^3 - 3t - a,$$

由 $f(1)>0$，$f(0)\leqslant 0$，$f(-\frac{1}{2})>0$，$f(-1)<0$，可知该方程有三个相异实根，而且 x 是唯一的正根。

由此可见，三等分角的问题就化为：给定长度为 a 的线段；求作一个长度为 x 的线段，而 x 满足 $4x^3-3x-a=0$。

从而，问题在于：利用直尺与圆规，什么样的长度是能够作得出来的，什么样的长度是作不出来的？为此，定义数集（给定长度单位）

$E=\{r$是实数：对给定长度a，可用直尺与圆规作出长度为$|r|$的线段$\}$

众所周知，E包含全部有理数，而且$\sqrt{2}$，$\sqrt{3}$，$\frac{\sqrt{5}-1}{2}$等都属于E。此外，当$a,b\in E$时，则$a\pm b,ab$，$\frac{a}{b}(b\neq 0)\in E$。更重要的是，若正数$a\in E$，则1与$a$的等比中项$\sqrt{a}\in E$。从而当$a,b\in E$且$a^2-4b>0$时，则$x^2-ax+b=0$的根

$$x=\frac{a\pm\sqrt{a^2-4b}}{2}$$

也属于E。下面引进抽象代数中的几个概念。

定义1 设U是复数的一个至少含有两个元素的子集，而且对U中任两个元素做加、减、乘、除（0不能做除数）运算的结果仍在U内，则称U是一个域。

显见，全体有理数组成的集合Q，全体实数组成的集合R，全体复数组成的集合C，都构成域．E也构成域（为什么）．我们可以这样来构造新的域。设U是一个域，u是给定的复数。令

$$U(u)=\left\{\frac{f(u)}{g(u)}:f,g\text{是系数属于}U\text{的一元多项式，}g(u)\neq 0\right\}.$$

易证$U(u)$是域．一般的，对给定的复数$u_1,\cdots,u_n$，令

$$U(u_1,\cdots,u_n)=\left\{\frac{f(u_1,\cdots,u_n)}{g(u_1,\cdots,u_n)}:f,g\text{是系数属于}U\text{的}\ n\text{元多项式，}g(u_1,\cdots,u_n)\neq 0\right\}.$$

易证$U(u_1,\cdots,u_n)$是域（证明留给读者）．特别的可取$U=Q$.

应用直尺、圆规作图，无非是保证能够作出直线$ax+by=c$以及圆$(x-a_1)^2+(y-b_1)^2=c_1^2$，其中$a,b,c,a_1,b_1,c_1$都是已经给定的线段的长度。当然，我们还可以使直线与直线，直线与圆，圆与圆相交。此外，就无所作为了。而它们相交的交点坐标若记为(x,y)，则又无非是

$$x\in Q(a,b,c,a_1,b_1,c_1)$$

或

$$x\in Q(a,b,c,a_1,b_1,c_1,d)$$

$$d^2\in Q(a,b,c,a_1,b_1,c_1)$$

等等。

在把几何问题化为代数问题后，现在可以阐明为什么三等分一角问题是不可解的了。其实，我们要证明的就是：若x满足$4x^3-3x^2-a=0$，则$x\notin E$。

为此，再引进两个概念：

定义2 设U是一个域，称n个复数$x_1,\cdots,x_n$在U上是线性相关的，如果存在U中不全为0的u_i（$i=1,2,\cdots,n$），使得

$$u_1x_1+u_2x_2+\cdots+u_nx_n=0$$

成立。否则，称$x_1,x_2,\cdots,x_n$在U上线性无关。

例如 1与$\sqrt{2}$在Q上线性无关，但1与$\sqrt{2}$在$Q(\sqrt{2})$上则线性相关。

定义3 若U与V都是域，而且$U\subset V$，如果V中存在n个数$x_1,x_2,\cdots,x_n$，它们在U上线性无关，而且

$$V = Ux_1 + \cdots + Ux_n$$
$$= \{u_1x_1 + \cdots + u_nx_n;\ u_i \in U\},$$

则称V关于U的维数为n，记为$[V:U]=n$。

例如 $[Q(\sqrt{2}):Q]=2$。

利用近世代数方法，已经证明

(i) 若$u\in E$，则$[Q(a,u):Q(a)]=2^m$，m是非负整数。

(ii) $[Q(a,x):Q(a)]=3$，如果$4x^3-3x-a=0$，而且$4t^3-3t-a$是$Q(a)$上的不可约多项式（即不能分为系数属于$Q(a)$的两个非常数多项式之积）。

由此可知，$x\overline{\in}E$的充分且必要条件是：多项式$4t^3-3t-a$是$Q(a)$上的不可约多项式。可以证明，当$\theta=\frac{\pi}{6}$，$\frac{\pi}{3}$或不为0的代数数时，$4t^3-3t-a$是$Q(a)$上的不可约多项式。

于是，三等分任意一角是不可解的。

粗略地说，我们把E的一部分看成是一个域$Q(a)$上的“向量空间”，这就允许去探讨它的“维数”，它就是所需要的不变量。上面指出，这些“向量空间”的“维数”总是2的方幂，它可能是4,8,16,…等等，这正是反映出下述事实：从给定的a出发，用直尺和圆规作图可以是很复杂的，这种复杂性使其所生成的数域在数域$Q(a)$上具有很高的维数。然而，对于大多数的角度，其三等分被证明为要求维数是3或3的倍数，而2的方幂不能同时又是3的倍数，这一矛盾导致其不可解性。

也许有人会问，为什么要限制作图工具只能是直尺与圆规？一种回答是：这些工具是最简单而不需要加工的，并且由于该命题的一时无法求解而被人们的好奇性与求知欲所推

动、继承。

这里给读者介绍两本极好的书：F.Klein的著作[2]立足于分析和代数，对几何的三大问题的不可解性均有论述；H. Eves[3]的著作则着重几何特征来论述。

顺便提醒读者的是，三等分角问题，方圆问题以及倍立方问题，虽然号称几何三大问题，但到了十九世纪，已失去原有的光彩。尤其是三等分角问题的解决，并没有给数学家带来太大的好处。几何三大问题根本不是几何学研究的主要课题。在十九世纪的几何学舞台上，还出现过不少有关平面三角形与圆的古怪结果，如九点圆定理等，今天我们有许多人几乎都不知道，在课堂上也不教这些内容，那是因为这些"美妙"的定理并没有使我们对几何概念的理解更加深刻，也无助于我们引发更加清晰和有启发性的直观想象力，充其量只不过可以满足人们的好奇心，就象谜语一样。然而，在成熟的解析几何面前，这种谜语不猜亦可。

例 6（n次根的线性组合） 每一个学生都学过整数的n次根，如$\sqrt{2}$，$\sqrt[3]{4}$等等。除了一些平凡的情形(如$\sqrt{4}$)外，它们都是无理数。然而，要判断一些n次根的线性组合，如

$$\sqrt[4]{3}+\sqrt[5]{4}+\sqrt[6]{72} \qquad (*)$$

是否无理数要困难得多！这里，人们极易想到，如果在(*)中的各项不是"明显地"被消去时，其和应为无理数。这一猜测是成立的。

首先，说一个正整数的n次根，我们指的是正实数。其次，为简明起见，我们来介绍这一结论的特殊情形：仅含有素因子2与3的整数的60次根（读者在希望作进一步推广时，将不会发生困难）。

定理 对于正实数集

$$\{\sqrt[60]{2^a3^b}: 0\leqslant a<60, 0\leqslant b<60\}$$

中3600个数在有理数域上是线性无关的。

注 读者也许不明白这个定理与“不可能性”有什么联系？为此，让我们选出其中四个数

$$1, \sqrt[4]{3}, \sqrt[5]{4}, \sqrt[6]{72}$$

来加以说明。如果不能找到不全为0的有理数a,b,c,d使得

$$a+b\sqrt[4]{3}+c\sqrt[5]{4}+d\sqrt[6]{72}=0.$$

（即线性无关），那末式(*)所表达的数就是无理数了。

令人惊奇的是，这一定理没有初等的证明（如果只限于考察平方根而不涉及立方根以及更高次方根，那末初等的解答是有的，见[4]。）

在一般形式下，[5]中用伽罗瓦理论证明了这一结果。

通过上述《不能证明的命题》的讨论，读者也许了解到“不变量”思想在其中的作用，有浅显的也有高层次的。我们可以这样说，这一思想及其老伙伴——变换的概念正处于许多近代数学分支的核心位置。

参 考 文 献

[1] H. Langman, Play Mathematics, Hafner, New York, 1962.

[2] F. Klein, Famous Problems of Elementary Geometry, repr. by Dover, New York, 1956.

[3] H. Eves, A Survey of Geometry, rev. ed., Allyn and Bacon, Boston, 1972.

[4] H. Flanders, (solution of a problem posed by D. J. Newman), (*Amer. Math. Monthly*, 67 (1960), 188—189.

[5] A. S. Besicovitch, On the linear independence of fractional powers of integers, *J. London Math. Soc.*, 15(1940), 3—6.

（周民强改译，潘承彪校）

第32届国际数学奥林匹克

竞 赛 试 题

1．设O是三角形ABC的内心，三内角$\angle BAC$，$\angle CBA$ $\angle ACB$的角平分线分别与其对边交于A',B',C'。证明：

$$\frac{1}{4}<\frac{AO\cdot BO\cdot CO}{AA'\cdot BB'\cdot CC'}\leqslant\frac{8}{27}.$$

2．设整数n大于6，$a_1,a_2,\cdots,a_k$是所有小于n且与n互素的正整数。如果

$$a_2-a_1=a_3-a_2=\cdots=a_k-a_{k-1}>0,$$

证明：n一定是素数或2的方幂。

3．设集合$S=\{1,2,3,\cdots,280\}$。求最小正整数n，使得S的每个有n个元素的子集都含有5个两两互素的数。

4．设G是一个有K条棱的连通图。证明：可以将G的棱标号$1,2,\cdots,K$，使得对属于两条或更多条棱的顶点，过该顶点的所有棱的标号的最大公约数是1。

5．设P是三角形ABC内的一点，证明：$\angle PAB$，$\angle PBC$，$\angle PCA$至少有一个不大于30°。

6．设$a>1$是给定的实数．试构造一个有界无穷数列x_0，$x_1,x_2,\cdots$，使得对任意的$i\neq j$有

$$|x_i-x_j|\geqslant|i-j|^{-a}.$$

第32届国际数学奥林匹克竞赛
试 题 解 答

潘承彪

国际数学奥林匹克竞赛是中学生的竞赛，因此，试题及其解答应在中学数学所能接受的范围之内，我们的解答将力求做到这一点。裘宗沪同志对本解答提出了宝贵意见，对此表示衷心感谢！

第1题的解答 作$AD, A'E$均垂直于BB'（见图1）．

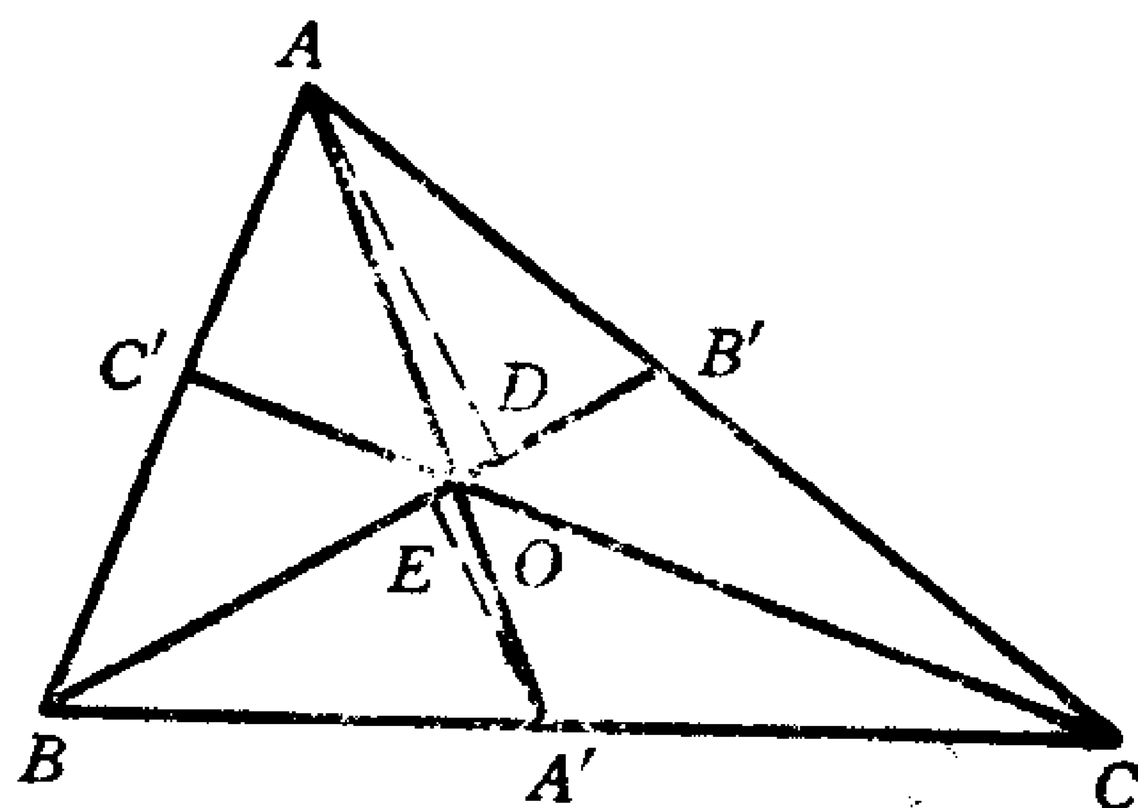


图 1

由$\triangle ABD\sim\triangle A'BE$推出$BA/BA'=AD/A'E$．由$\triangle A'OE\sim\triangle AOD$推出$AD/A'E=AO/A'O$．因而有

$$BA/BA'=AO/A'O.$$

同理可得

$$CA/CA' = AO/A'O.$$

由以上两式得到

$$(BA + CA)/BC = AO/A'O,$$

进而有

$$\alpha = (BA + CA)/(BA + CA + BC) = AO/AA'.$$

这就是平面几何中熟知的分角线定理。同理可得

$$\beta = (AB + CB)/(BA + CA + BC) = BO/BB',$$

$$\gamma = (AC + BC)/(BA + CA + BC) = CO/CC'.$$

因而有（为什么）

$$1/2 < \alpha < 1,\ 1/2 < \beta < 1,\ 1/2 < \gamma < 1;$$

$$\alpha + \beta + \gamma = 2.$$

利用三个非负数的几何平均值不超过它们的算术平均值，得到

$$(\alpha \cdot \beta \cdot \gamma)^{1/3} \leqslant (\alpha + \beta + \gamma)/3 = 2/3,$$

由此就推出所要证的右半不等式。

为证左半不等式，不妨设$BC \geqslant CA \geqslant AB$，即$\gamma \geqslant \beta \geqslant \alpha$。我们有

$$\alpha \cdot \beta \cdot \gamma = \alpha \cdot \beta \cdot (2 - \alpha - \beta)$$

$$= \beta \left[\left(\frac{2-\beta}{2}\right)^2 - \left(\frac{2-\beta}{2} - \alpha\right)^2\right].$$

由假设知$2 - \beta = \alpha + \gamma \geqslant 2\alpha$，由此及$\alpha > 1/2$从上式推得

$$\alpha \cdot \beta \cdot \gamma > \beta \left[\left(\frac{2-\beta}{2}\right)^2 - \left(\frac{1-\beta}{2}\right)^2\right]$$

$$= \frac{1}{2}\left[\frac{9}{16} - \left(\beta - \frac{3}{4}\right)^2\right].$$

利用$1/2<\beta<1$，从上式得

$$\alpha\cdot\beta\cdot\gamma>1/4.$$

这就证明了所要的左半不等式。

第 2 题的解答　记$d=a_2-a_1$。由条件知，$a_1=1$，$a_k=n-1$，$d\geqslant1$，及

$$a_{j+1}=1+jd,\ j=0,1,\cdots,k-1.$$

(1) a_2是素数，且若素数$p<a_2$，则$p|n$。由条件知a_2是大于1且与n互素的最小正整数，所以$a_2>1$不可能是合数(为什么)，即一定是素数。此外，由最小性推出素数$p<a_2$，p与n一定不互素，即$p|n$。

下面按$a_2=2$，$a_2=3$，$a_2>3$三种情形来讨论。

(2) $a_2=2$。这时$d=1$，$a_{j+1}=1+j$，$0\leqslant j\leqslant k-1$。由于$n-1=a_k=k$，所以$n$与小于$n$的所有正整数互素，因此，$n$必为素数（为什么）。

(3) $a_2=3$，这时$d=2$，$a_{j+1}=1+2j$，$0\leqslant j\leqslant k-1$。由于$n-1=a_k=1+2(k-1)$，所以$k=n/2$。这时$n$是偶数，且与所有小于它的正奇数互素。因此，除2外$n$不能有别的素因数，即$n=2^s$。

(4) $a_2>3$。这时由(1)知$3|n$。由此及$n-1=a_k=1+(k-1)d$推出$3\nmid d$。因此，$3|1+d$或$3|1+2d$有且仅有一个成立(为什么)。由于$a_2=1+d>3$是素数，所以必有$3\nmid1+d$，$3|1+2d$。因此得$k=2$，即与n互素的且不超过n的正整数仅有1和a_2两个。下面来证明：当$n>6$时必有$k>2$.所以这种情形是不可能出现的。为此设$n=2^r\cdot m>6$，$2\nmid m$。

(i) $r=0$.这时$m>6$.因此，1,2,4均与奇数$n=m$互素，因此有$k>2$；

(ii) $m=1$。这时$r\geqslant3$，$1,3,5,7,\cdots$均与$n=2^r$互素，故有$k>2$；

(iii) $m=3$。这时$r\geqslant2$，$1,5,7$必与$n=2^r\cdot3$互素，故有$k>2$；

(iv) $r\geqslant1$，$m\geqslant5$。这时$1,m-2,m+2<n$均与n互素，故有$k>2$。这就对所有可能的情形证明了所要的结论。应该指出：这里的k就是初等数论中的 Euler 函数$\phi(n)$，但是，用$\phi(n)$的表示式来证明这结论是超出了中学数学范围。

读者不难看出，只要假定$n\neq1,6$，本题就成立，而且n等于素数或2^s时，所有不超过n且与n互素的正整数按大小顺序排列为$a_1,a_2,\cdots,a_k$时，必满足题中的条件。

第3题的解答 怎样的整数集合M会使它的任意5个数一定不是两两互素？一个容易想到的条件是：存在四个不同的素数p_1,p_2,p_3,p_4，使M中的数至少被一个p_i整除。如何从S中找出满足这样性质的子集M，使得M有尽可能多的元素。只有考察了这一点，才可能对题中要求的最小正整数n有所了解。为使$M\subseteq S$按上述条件来构造，且有尽可能多的元素，显然应取尽可能小的p_1,p_2,p_3,p_4。现取$p_1=2,p_2=3,p_3=5,p_4=7$。设d是给定的正整数，记

$$S_d=\{s:s\in S,d\mid s\}.$$

取$M=S_2\cup S_3\cup S_5\cup S_7\subset S$。$M$中任意5个数一定有2个同属于某一个$S_d$，即一定不两两互素。下面用容斥原理来求集合$M$的元素个数(以$|A|$表集合$A$的元素个数)：

$$\begin{aligned}|M|=&(|S_2|+|S_3|+|S_5|+|S_7|)\\&-(|S_6|+|S_{10}|+|S_{14}|+|S_{15}|+|S_{21}|+|S_{35}|)\\&+(|S_{30}|+|S_{42}|+|S_{70}|+|S_{105}|)-|S_{210}|\end{aligned}$$

$$= (140+93+56+40)-(46+28+20+18+13+8)$$
$$+(9+6+4+2)-1$$
$$=216.$$

由此知必有 $n\geqslant 217$。

另一方面，如果 r 个整数集合 $B_1,\cdots,B_r$，它们两两不交，每个 B_i 的元素个数不少于 5 个，且每个 B_i 中的整数均两两互素。设 $B=B_1\cup B_2\cup\cdots\cup B_r$。显见，在 B 中任取 $4r+1$ 个元素必有 5 个整数属于某一个 B_i，因而必两两互素。如果这些 $B_i(1\leqslant i\leqslant r)$ 都是 S 的子集，那么，在 S 中任取 $|S|-|B|+4r+1$ 个正整数就必有 $4r+1$ 个属于 B，因而必有 5 个正整数两两互素。下面我们来构造 B_i。为使能从 S 中取尽可能少的数达到题中的要求，就需要 $|B|$ 尽可能大而 r 尽可能小。为此取

$$B_1=\{s:s\in S, s=1 \text{ 或素数}\},$$
$$B_2=\{2^2,3^2,5^2,7^2,11^2,13^2\},$$
$$B_3=\{2\cdot 139,3\cdot 89,5\cdot 53,7\cdot 37,11\cdot 23,13\cdot 19\},$$
$$B_4=\{2\cdot 137,3\cdot 83,5\cdot 47,7\cdot 31,11\cdot 19,13\cdot 17\},$$
$$B_5=\{2\cdot 131,3\cdot 79,5\cdot 43,7\cdot 29,11\cdot 17\},$$
$$B_6=\{2\cdot 127,3\cdot 73,5\cdot 41,7\cdot 23,11\cdot 13\}.$$

容易验证这些集合满足以上所说的要求。$r=6$，

$$|B|=|B_1|+\cdots+|B_6|=60+6+6+6+5+5=88.$$

因此，在 S 中任取 $280-88+4\cdot 6+1=217$ 个数，必有 5 个是两两互素的。所以 $n\leqslant 217$。

综合以上两部分即得 $n=217$。

第 4 题的解答 用以下方法对棱标号就可满足题中的要求。任取一顶点记作 v_0，由于是连通图，v_0 必至少属于一条

棱。从v_0出发沿G中的棱不重复地前进，即已通过的棱不允许再次通过，但顶点允许多次经过，直到不能再这样前进为止。依次记已通过的顶点为$v_0,v_1,\cdots,v_{l_1}$（注意不同的标号可能对应同一个顶点，但相邻的v_i,v_{i+1}不会是同一个顶点），连结顶点$v_{i-1},v_i(1\leqslant i\leqslant l_1)$，并给以棱标号$i$。这样就有$l_1$条棱（它们是两两不同的）标号为$1,2,\cdots,l_1$。显见$1\leqslant l_1\leqslant K$。在已通过的顶点中，可能除了$v_0$和$v_{l_1}$外，其余每个顶点至少属于两条已被标号的棱，其标号数中必有两个是相邻正整数。顶点v_0必属于标号为1的棱。顶点v_{l_1}要末也至少属于两条已被标号的棱，其标号数中必有两个是相邻正整数，要末它在连通图G中只属于一条棱（为什么）。若$l_1=K$，则所有棱均已标号，由上述说明知标号满足要求。

若$1\leqslant l_1<K$。由图的连通性知，在顶点$v_0,v_1,\cdots,v_{l_1-1}$中必有顶点属于还未标号的棱。任取这样一个顶点，记为v_{l_1+1}。从这一顶点出发，按上述规则，沿G中未被通过的棱（即未标号的棱）不重复地前进，直到不能继续前进为止。依次记通过的顶点为$v_{l_1+1},v_{l_1+2},\cdots,v_{l_1+l_2-1}$，连结$v_i$，$v_{i+1}(l_1+1\leqslant i\leqslant l_1+l_2)$的棱标号$i$。这样，就又有不同的$l_2$条棱标号为$l_1+1,\cdots,l_1+l_2,1\leqslant l_2\leqslant K-l_1$。若$l_1+l_2<K$，则继续用这样的方法标号。由于总共只有$K$条棱，所以，最后一定将所有的棱标号为$1,2,\cdots,K$。

由我们的标号方法知，对G的任一顶点如果它至少属于两条棱，那么，当它为v_0时，必有过该点的棱标号为1；当它不是v_0时，必有过该点的两条棱标号为相邻正整数。因此，过这样的顶点的所有棱的标号数的最大公约数必为1。图2给出了有11条棱的连通图，$l_1=8,l_2=3$。图3给出的不

连通图，无论怎样标号都不会满足题目的要求(为什么)。本题不是初等数论题。

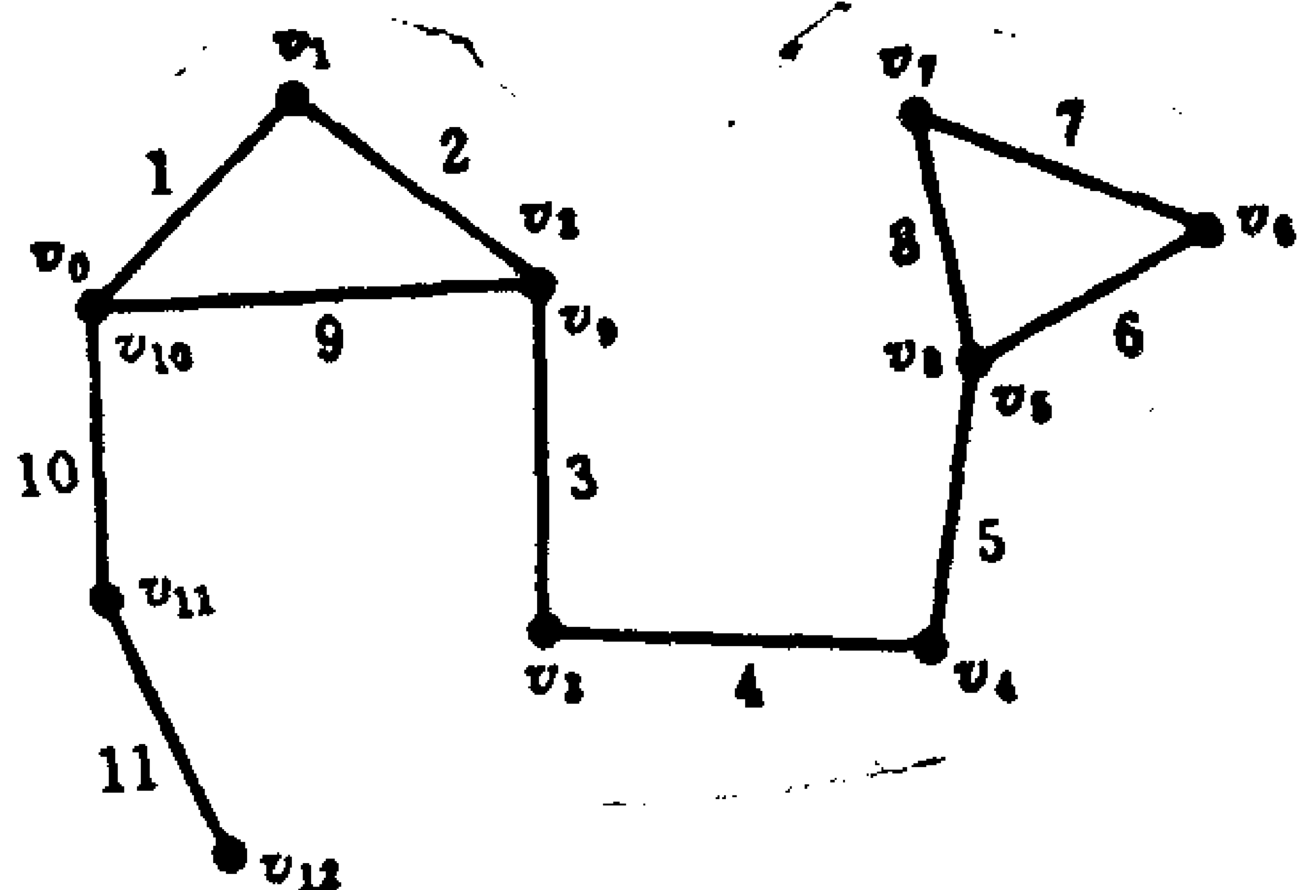


图 2

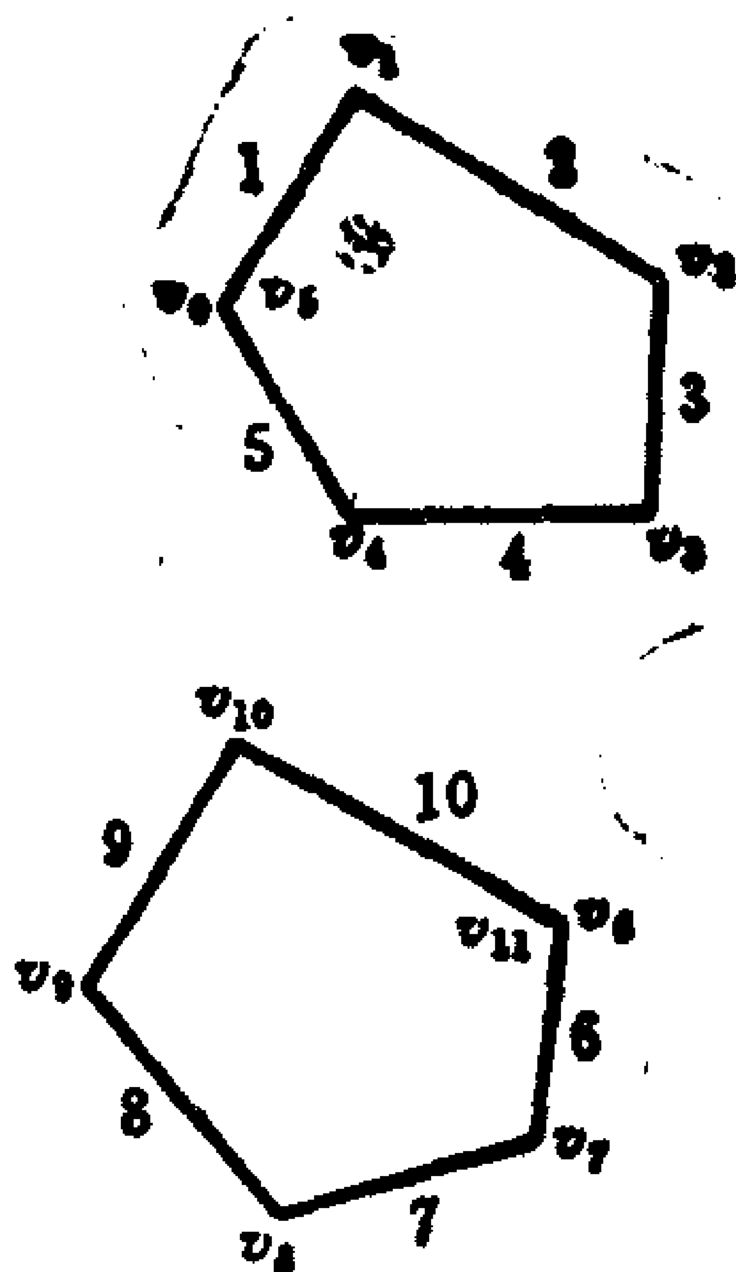


图 3

第 5 题的解答 记 $\alpha=\angle CAB, \alpha_1=\angle PAB, \beta=\angle ABC, \beta_1=\angle PBC, \gamma=\angle BCA, \gamma_1=\angle PCA$(图 4)。

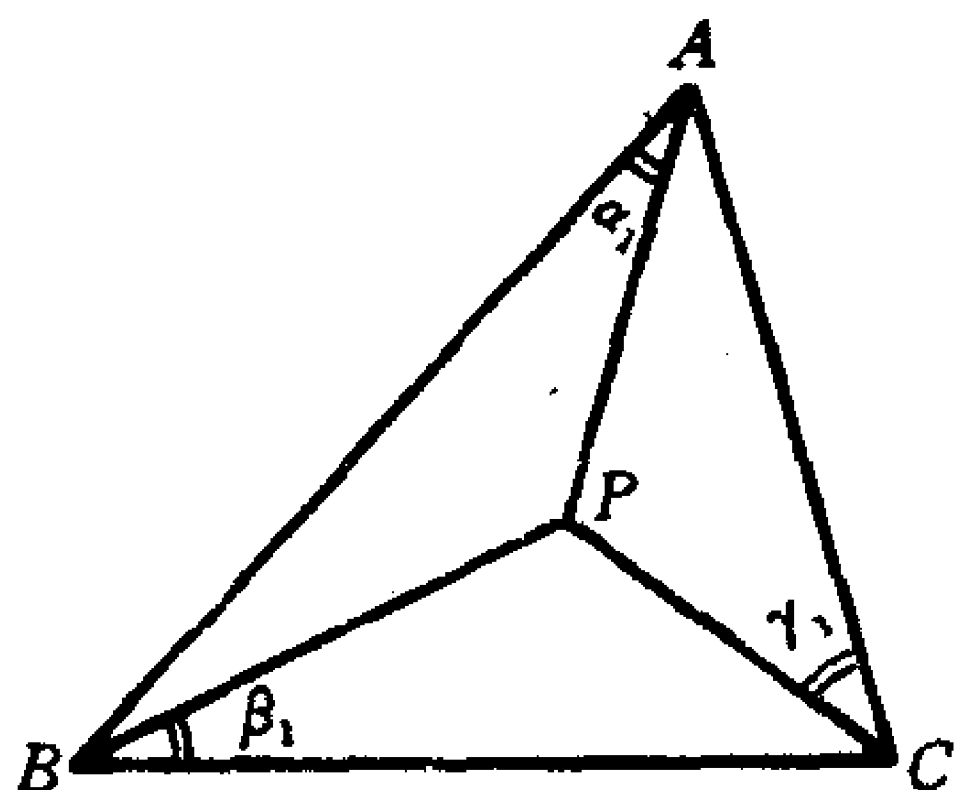


图 4

我们有

$$PA\ \sin\alpha_1=PB\sin(\beta-\beta_1),$$
$$PB\sin\beta_1=PC\sin(\gamma-\gamma_1),$$
$$PC\sin\gamma_1=PA\sin(\alpha-\alpha_1).$$

由以上三式即得

$$\sin\alpha_1\cdot\sin\beta_1\cdot\sin\gamma_1=\sin(\alpha-\alpha_1)\cdot\sin(\beta-\beta_1)\cdot\sin(\gamma-\gamma_1).$$

进而有

$$\sin^2\alpha_1\cdot\sin^2\beta_1\cdot\sin^2\gamma_1=\sin\alpha_1\cdot\sin(\alpha-\alpha_1)\cdot\sin\beta_1\cdot\sin(\beta-\beta_1)\times\sin\gamma_1\cdot\sin(\gamma-\gamma_1).$$

由熟知的不等式得（注意 $0\leqslant\alpha_1\leqslant\alpha<\pi$，$|\alpha-2\alpha_1|<\pi$）

$$(\sin\alpha_1\cdot\sin(\alpha-\alpha_1))^{1/2}\leqslant\frac{\sin\alpha_1+\sin(\alpha-\alpha_1)}{2}$$

$$=\sin\frac{\alpha}{2}\cdot\cos\frac{\alpha-2\alpha_1}{2}\leqslant\sin\frac{\alpha}{2}.$$

等号当且仅当 $\alpha_1=\alpha/2$ 时成立。同理可得

$$(\sin\beta_1\cdot\sin(\beta-\beta_1))^{1/2}\leqslant\sin\beta/2,$$

$$(\sin\gamma_1\cdot\sin(\gamma-\gamma_1))^{1/2}\leqslant\sin\gamma/2.$$

等号分别当且仅当 $\beta_1=\beta/2$，$\gamma_1=\gamma/2$ 时成立。因而有

$$\sin\alpha_1\cdot\sin\beta_1\cdot\sin\gamma_1\leqslant\sin\alpha/2\cdot\sin\beta/2\cdot\sin\gamma/2. \qquad (1)$$

再利用三个非负数的几何平均数不大于它们的算术平均数，得到

$$(\sin\alpha/2\cdot\sin\beta/2\cdot\sin\gamma/2)^{1/3}\leqslant(\sin\alpha/2+\sin\beta/2+\sin\gamma/2)/3. \qquad (2)$$

以下不妨设 $\alpha\leqslant\beta\leqslant\gamma$。我们有

$$\sin\frac{\alpha}{2}+\sin\frac{\beta}{2}+\sin\frac{\gamma}{2}=2\sin\frac{(\beta+\alpha)}{4}\cdot\cos\frac{(\beta-\alpha)}{4}+\sin\frac{\gamma}{2}$$

$$\leqslant2\sin(\beta+\alpha)/4+\sin\gamma/2, \qquad (3)$$

等号当且仅当 $\alpha=\beta$ 时成立。注意到 $\alpha+\beta+\gamma=\pi$，及假定 $\alpha\leqslant\beta\leqslant\gamma$，可设

$$\gamma/2=\pi/6+2\theta,\quad (\beta+\alpha)/4=\pi/6-\theta,\quad 0\leqslant\theta<\pi/6,$$

得到

$$2\sin(\beta+\alpha)/4+\sin\gamma/2=2\sin(\pi/6-\theta)+\sin(\pi/6+2\theta)$$

$$=(2\cos\theta+\cos2\theta)\sin\pi/6+(-2\sin\theta+\sin2\theta)\cos\pi/6$$

$$=(2\cos\theta+\cos2\theta)\sin\pi/6+2(-1+\cos\theta)\sin\theta\cdot\cos\pi/6$$

$\leqslant 3/2$. (4)

等号仅当 $\theta=0$ 时成立。由式(2),(3),(4)推出

$$\sin\alpha/2+\sin\beta/2+\sin\gamma/2\leqslant 3/2, \quad (5)$$

$$\sin\alpha/2\cdot\sin\beta/2\cdot\sin\gamma/2\leqslant 1/8, \quad (6)$$

等号仅当 $\alpha=\beta=\gamma=\pi/3$ 时成立。由式(6)和(1)即得

$$\sin\alpha_1\cdot\sin\beta_1\cdot\sin\gamma_1\leqslant 1/8.$$

由此推出左边三项至少有一个$\leqslant 1/2$，不妨设 $\sin\alpha_1\leqslant 1/2$。这样，必有 $\alpha_1\leqslant 30°$，或 $\alpha_1\geqslant 150°$，在后一情形，β_1,γ_1 均小于 30°。证毕。

不等式(5)和(6)是两个很著名的有关三角形的内角的不等式。这里给出了与通常不同的证明。(参看：O. Bottema, 几何不等式，§2，北京大学出版社，1991)。

第 6 题的解答　下面给出两个解法。

解答一　由题的要求知，所构造的数列 x_i 应和足标 i 有关，以保证两数之差的绝对值 $|x_i-x_j|$ $(i\neq j)$ 相对于足标之差 $|i-j|$ 不是太小。一个方法是通过足标的 k 进制表示来给出数列 x_i。下面利用二进位表示。设 i 是非负整数，它的二进位表示是

$$i=b_0+b_1\cdot 2+b_2\cdot 2^2+\cdots+b_r\cdot 2^r,$$

其中 $b_0,b_1,\cdots,b_r$ 取值 0 或 1。令

$$y_i=b_0+b_1\cdot 2^{-a}+b_2\cdot 2^{-2a}+\cdots+b_r\cdot 2^{-ra},$$

$$i=0,1,2,\cdots.$$

显有 $|y_i|<(1-2^{-a})^{-1}$，所以是一有界数列。下面来估计 $|y_i-y_j|, j\neq i$。设

$$j=c_0+c_1\cdot 2+c_2\cdot 2^2+\cdots+c_s\cdot 2^s.$$

由于 $j\neq i$，一定有非负整数 t_0 使得

$$b_{t_0} \neq c_{t_0}, \quad b_t = c_t, \quad 0 \leqslant t < t_0.$$

因而有$|i-j| \geqslant 2^{t_0}$(为什么)，以及

$$|y_i - y_j| > 2^{-t_0 a} - (2^{-(t_0+1)a} + 2^{-(t_0+2)a} + \cdots)$$

$$= 2^{-t_0 a} \cdot \frac{2^a - 2}{2^a - 1} \geqslant |i-j|^{-a} \cdot \frac{2^a - 2}{2^a - 1}.$$

由上式知，取数列

$$x_i = \frac{2^a - 2}{2^a - 1} y_i, \quad i \geqslant 0,$$

即满足要求。

这里要指出两点：一是用这样的方法构造的数列一定要求题中的 $a > 1$；二是这是一个有理数列。

解答二 这个解答本质上是由我国选手罗炜提出的，而且允许把问题改进为$a=1$。他的方法是基于丢番图逼近理论中的这样一个结论：对任给的实二次无理数 α(即实数 α 不是有理数，且是某个二次整系数多项式的根)，一定存在一个正常数 c (可以和 α 有关)，使得对任意整数 $p \neq 0$ 及 q，必有

$$|p\alpha - q| \geqslant c/|p|, \tag{7}$$

即

$$|\alpha - q/p| \geqslant c/p^2. \tag{8}$$

这表明这样的无理数 α 用有理数 q/p 来逼近时，不可能逼近得太好(关于这方面的初步知识可参看：R.P.布恩，数论入门，第十一章，高等教育出版社，1990。这是一本很好的初等数论入门书)。

我们先来证明：若式(7)对某个实二次无理数 α 成立，就可推出本题当 $a=1$ 时成立。取数列

$$x_i = c^{-1}(a\cdot i-[a\cdot i]),\quad i=0,1,2,\cdots,$$

这里$[t]$表不超过实数 t 的最大整数。显有 $0\leqslant x_i<c^{-1}$，所以是有界数列。若式(7)成立，则有

$$|x_i-x_j|=c^{-1}|a\cdot(i-j)-([a\cdot i]-[a\cdot j])|$$
$$\geqslant|i-j|^{-1},\quad i\neq j.$$

这就推出有界无穷数列 x_i 满足本题当 $a=1$ 时的要求。

下面来证明：当 $a=\sqrt{2}, c=(2+\sqrt{2})^{-1}$ 时，对任意整数 $p\neq0$ 及 q，式(7)(即(8))成立。先假定

$$1\leqslant q/p\leqslant2. \tag{9}$$

由 $2p^2-q^2\neq0$(为什么)推出：

$$1\leqslant|2p^2-q^2| = |p\sqrt{2}+q||p\sqrt{2}-q|$$
$$=|p||\sqrt{2}+q/p||p\sqrt{2}-q|$$
$$\leqslant(2+\sqrt{2})|p||p\sqrt{2}-q|,$$

最后一步用到了条件(9)。当 $q/p<1$ 时，显有

$$|\sqrt{2}-q/p|>\sqrt{2}-1>(2+\sqrt{2})^{-1}$$
$$\geqslant(2+\sqrt{2})^{-1}/p^2;$$

当 $q/p>2$ 时，

$$|\sqrt{2}-q/p|>2-\sqrt{2}>(2+\sqrt{2})^{-1}$$
$$\geqslant(2+\sqrt{2})^{-1}/p^2.$$

由以上三式就推出所要的结论。因此，我们可取

$$x_i=(2+\sqrt{2})(\sqrt{2}\cdot i-[\sqrt{2}\cdot i]),$$
$$i=0,1,2,\cdots.$$

应该指出，这里构造的数列$\{x_i\}$是无理数列，而在解答一中构造的是有理数列。对 $a=1$ 是否也能构造出一个满足条件的有理数列呢？我还不知道。可能回答是否定的。

初等数学问题(2)解答

1．由费马定理① 知：$p^4\equiv1 \pmod 5$。

$$p^4-1=(p^2+1)(p^2-1)\equiv0 \pmod 5。\qquad (1)$$

因为p是奇数，所以p^2+1与p^2-1均为偶数，

$$2|p^2+1,\quad 2|p^2-1,$$

又$(2,5)=1$。结合(1)知$10|p^2+1$或$10|p^2-1$。

2．证明　因为n是不小于3的奇数，所以可设$n=2k+1(k\in N)$。又因为$2^6=64\equiv1 \pmod 9$，将k按$3j,3j+1,3j+2(j\in N)$分类有：

(i) $k=3j$时，

$$2^{n-1}(2^n-1)=2^{6j}(2^{6j+1}-1)\equiv2^0(2-1)\equiv1 \pmod 9;$$

(ii) $k=3j+1$时，

$$2^{n-1}(2^n-1)=2^{6j+2}(2^{6j+3}-1)\equiv2^2(2^3-1)=28\equiv1 \pmod 9;$$

(iii) $k=3j+2$时，

$$2^{n-1}(2^n-1)=2^{6j+4}(2^{6j+5}-1)$$
$$\equiv2^4(2^5-1)\equiv7\times4\equiv1 \pmod 9。$$

所以$2^{n-1}(2^n-1)\equiv1 \pmod 9$对任何不小于3的奇数$n$成立。

3．注意到所给数列的前14项中每项均可分解为两个素

① 费马定理：若p是素数，a不能被p整除，则$a^{p-1}\equiv1 \pmod p$。

数的乘积，且数列是单调上升的。故该数列第15、16、17项是39、46、49。

4. 证明　因为n是素数，$n\nmid r$，由费马定理知：$r^{n-1}\equiv 1(\bmod n)$，所以

$$\frac{r^{n-1}-1}{n}\in \boldsymbol{Z}.$$

由此可构造如下恒等式：

$$\left(2^{\frac{r^{n-1}-1}{n}}\right)^n+\left(2^{\frac{r^{n-1}-1}{n}}\right)^n=\left(2^{r^{n-2}}\right)^r.$$

这说明不定方程 $x^n+y^n=z^r$ 至少有如下整数解

$$\left(2^{\frac{r^{n-1}-1}{n}},2^{\frac{r^{n-1}-1}{n}},2^{r^{n-2}}\right).$$

5.① 证明　由条件 $x_i>0$，$\sum_{i=1}^{n}x_i=1$，$x_{n+1}=x_1(n>6)$和算术-几何平均值不等式知：

$$\prod_{i=1}^{n}(x_i+x_{i+1})\leqslant\left[\frac{\sum_{i=1}^{n}(x_i+x_{i+1})}{n}\right]^n$$

$$=\left[\frac{2\sum_{i=1}^{n}x_i}{n}\right]^n=\left(\frac{2}{n}\right)^n,$$

所以

---

① 原题结论有错。要证的结论应为：$\prod_{i=1}^{n}(x_i+x_{i+1})^{-1}\geqslant n!$。——编者注

$$\prod_{i=1}^{n}(x_i+x_{i+1})^{-1}\geqslant\left(\frac{n}{2}\right)^n.$$

下面证明$\left(\frac{n}{2}\right)^n>n!$ $(n>6)$.

当 $n=7$ 时，有

$6\times1\times2^3<7^2$，　$5\times4\times2<7^2$，　$3\times2\times2^3<7^2$，

以上三式左、右乘之得：

$$6!\cdot2^7<7^6,$$

所以

$$7!\cdot2^7<7^7.$$

假设 $n=k>6$ 时，有　$k!\cdot2^k<k^k$.

当 $n=k+1$ 时，由于$(k+1)!\cdot2^{k+1}=k!\cdot2^k\cdot2(k+1)$，由归纳假设：

$$k!\cdot2^k\cdot2(k+1)<k^k\cdot2\cdot(k+1),$$

因为

$$2<\left(1+\frac{1}{k}\right)^k\quad(k>1),\tag{2}$$

所以

$$2\cdot k^k<(k+1)^k,$$

结合(2)知

$$(k+1)!\cdot2^{k+1}<(k+1)^{k+1}.$$

由数学归纳法原理对 $n>6(n\in\boldsymbol{N})$有 $2^n\cdot n!<n^n$，即

$$\left(\frac{n}{2}\right)^n>n!,$$

这就证明了

$$\prod_{i=1}^{n}(x_i+x_{i+1})^{-1}\geqslant\left(\frac{n}{2}\right)^n>n!\quad(n>6).$$

6. 证明 （反证） 假设 $\lim\limits_{n\to\infty}\operatorname{tg} n=p<\infty, p\in \boldsymbol{R}$。因为

$$(1-\operatorname{tg} n\cdot\operatorname{tg}1)\operatorname{tg}(n-1)=\operatorname{tg} n-\operatorname{tg}1, \qquad (3)$$

对(3)令 $n\to+\infty$，知有

$$(1-p\cdot\operatorname{tg}1)\cdot p=p-\operatorname{tg}1.$$

由此得出 $p^2=-1$，这与 $p\in \boldsymbol{R}$ 矛盾。所以$\{\operatorname{tg} n\}$是发散的。

7. 证明 由伯努利不等式：$(1+m)^a\geqslant 1+am (a>1, m>-1)$。取 $a=n>1$； $m=-\dfrac{1}{n^2}>-1$。有

$$\left(1-\frac{1}{n^2}\right)^n\geqslant 1-\frac{1}{n}=\frac{n-1}{n},$$

$$\frac{n+1}{n-1}\left(1-\frac{1}{n^2}\right)^n\geqslant\frac{n+1}{n-1}\cdot\frac{n-1}{n}=\frac{n+1}{n}>1,$$

所以

$$\frac{(n+1)(n^2-1)^n}{(n-1)\cdot n^{2n}}>1,$$

$$\frac{(n+1)^{n+1}(n-1)^{n-1}}{n^{2n}}>1, \quad n=2,3,\cdots,$$

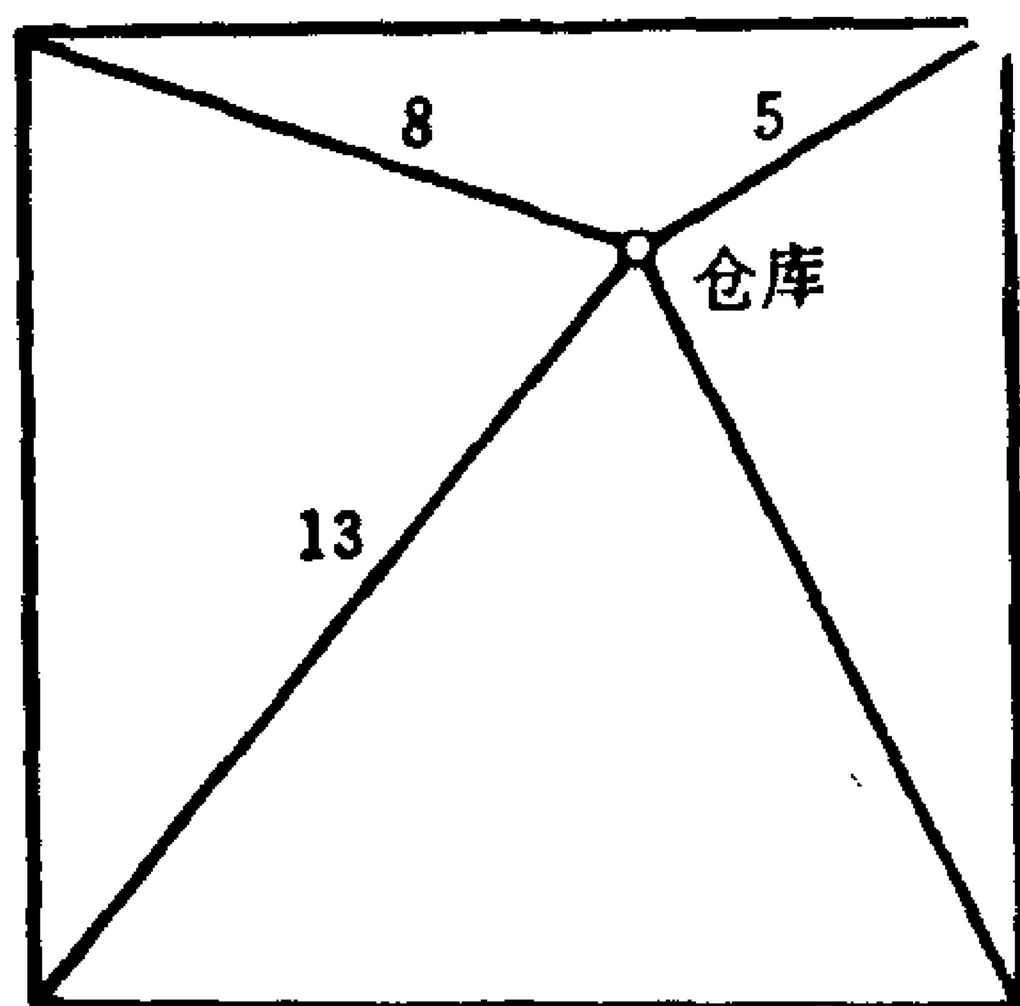
即

$$\frac{(n+1)^{n+1}}{n^n}>\frac{n^n}{(n-1)^{n-1}}.$$

（张恩明提供，陈剑刚校）

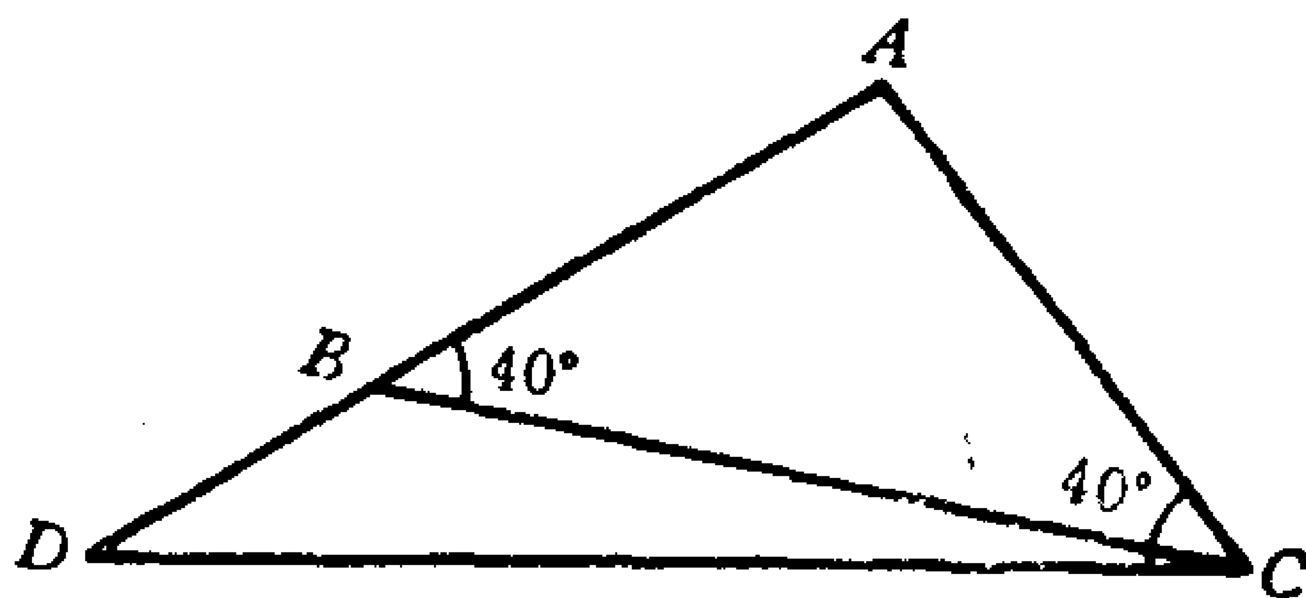
初等数学问题[①]（3）

1．一块正方形的地域以四条马路为界，仓库位置距西南 13km，西北角 8km，东北角 5km，问仓库距最近的马路有多少公里？



（第 1 题图）

2．已知：锐角△ABC 的三条高线分别是 AD，BE，CF。证明：这三条高线是△DEF 的内角平分线。



（第 3 题图）

---

① 这个专栏是由北京大学附中陈剑刚主持下编写的，每期都由具有丰富的中学数学教学经验的教师给出若干有趣味性的数学问题，在下一册给出这些问题的解答。

3．$\triangle ABC$中，$\angle ABC=\angle ACB=40°$，延长$AB$到$D$，使得$AD=BC$，求$\angle BCD$的大小。

4．平面几何中有一个著名定理（Lehmus）：如果一个三角形的两条内角平分线长是相等的，则该三角形是等腰的。现请你证明：若一个三角形的两条外角平分线长相等，则该三角形不必须是等腰的。

5．已知：$O\text{-}ABCDE$是一个正五棱锥，$\angle AOB=60°$，求$\angle AOC$。

（张思明选编，陈剑刚校）

数学小丛书——智慧之花

(4)

主 要 目 录

[illegible]

[illegible]

[illegible]
“n[illegible]”

[illegible]

Euclid[illegible]

[illegible]

[illegible]3:4:5[illegible]

[illegible]

[illegible]Euclid[illegible]

[illegible]

[illegible]

[illegible]

[illegible]2[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]32[illegible]

[illegible]32[illegible]

[illegible]2[illegible]

[illegible]3[illegible]